

Thu Nov 7 09:13:10 2002

Db 1 MDLEGRNGGAKKNFFKLNKSEK-DKKEKKPTVSFMSFRYSNWLDKLYMVVGTAAI 59
Qy 60 IHGAALPLMLVFGNMTDFANAGISRNKTPPVIIINESITNNQHTHINLEEMTYAY 119
Db 60 IHGAGLPLMLVFGEMTDIFANAG-NLEDLSNITNRSNDINDGFFMN-LEEDMTYAY 117
Qy 120 YSGIGAGVLVAAYIOVSFWCLAAAGROILKIRKQFFHAIMRQEIOWFDVHVGELNRLTD 179
Db 118 YSGIGAGVLVAAYIOVSFWCLAAAGROILKIRKQFFHAIMRQEIOWFDVHVGELNRLTD 177
Qy 180 DYSKINEGIDKIGMFFHSIATFTFTGTVGTRGKMLTLVLAISPVGLSAAIWAKILS 239
Db 178 DVSQKINEGIDKIGMFFHSIATFTFTGTVGTRGKMLTLVLAISPVGLSAAIWAKILS 237
Qy 240 SFTDKELLAYAKAGAAVEVLAAITVIAFGQKKELRYNKNLEAKGIGIKKAITANI 299
Db 238 SFTDKELLAYAKAGAAVEVLAAITVIAFGQKKELRYNKNLEAKGIGIKKAITANI 297
Qy 300 SIGAFLLIYASALAFWGTGTVLSSEYSIGQVLTFFSVLIGAFSIGQASPSIEAFAN 359
Db 298 SIGAFLLIYASALAFWGTGTVLSSEYSIGQVLTFFSVLIGAFSIGQASPSIEAFAN 357
Qy 360 ARGAAEYIFKIIDNKPISIDYSKSGHKPDNIKNLEFNHVSFYSRKEVKILKGLNLKV 419
Db 358 ARGAAEYIFKIIDNKPISIDYSKSGHKPDNIKNLEFNHVSFYSRKEVKILKGLNLKV 417
Qy 420 QSGOTVALVGNCGGKSTTVQLMQRLYDPTGVMVSDGQDRTINVRHLREITGVVSOEP 479
Db 418 QSGOTVALVGNCGGKSTTVQLMQRLYDPTGVMVSDGQDRTINVRHLREITGVVSOEP 477
Qy 480 VLFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPNKFDTLVGERGQALSGGOK 539
Db 478 VLFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPNKFDTLVGERGQALSGGOK 537

Qy 1140 RVVSHHEIMQAQKANTHHFIETLPEKYNTRVGDKGTOLSGGQKQRIATARALVROPHIL 1199
Db 1138 RVVSHHEIMQAQKANTHHFIETLPEKYNTRVGDKGTOLSGGQKQRIATARALVROPHIL 1197
Qy 1200 LDEATSAIDTESKVVQEAALDKAREGRTCTIVIAHRLSTONADLIYVFGNGKVEGHTH 1259
Db 1198 LDEATSAIDTESKVVQEAALDKAREGRTCTIVIAHRLSTONADLIYVFGNGKVEGHTH 1257
Qy 1260 QQLLAQKGIYFSMVSQVQAGKR 1281
Db 1258 QQLLAQKGIYFSMVSQVQAGTKR 1279
RESULT 2
5206352-4
; Patent No. 5206352
; Applicant: Roninon, Igor B.; Pastan Ira H.: Gottesman,
; Michael M.
; TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
; SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/622,836
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 892,575
; FILING DATE: 01-AUG-1986
; APPLICATION NUMBER: 845,610
; FILING DATE: 28-MAR-1986
; SEQ ID NO: 4:
; LENGTH: 1280
5206352-4
Query Match 90.3%; Score 5849; DB 6; Length 1280;
Best Local Similarity 90.6%; Pred. No. 0;
Matches 1161; Conservative 56; Mismatches 61; Indels 4; Gaps 4;
Qy 1 MDPEGRNGGAKKNFFKLNKSEK-DKKEKKPTVSFMSFRYSNWLDKLYMVVGTAAI 59
Db 1 MDLEGRNGGAKKNFFKLNKSEK-DKKEKKPTVSFMSFRYSNWLDKLYMVVGTAAI 59
Qy 60 IHGAALPLMLVFGNMTDFANAGISRNKTPPVIIINESITNNQHTHINLEEMTYAY 119
Db 60 IHGAGLPLMLVFGEMTDIFANAG-NLEDLSNITNRSNDINDGFFMN-LEEDMTYAY 117
Qy 120 YSGIGAGVLVAAYIOVSFWCLAAAGROILKIRKQFFHAIMRQEIOWFDVHVGELNRLTD 179
Db 118 YSGIGAGVLVAAYIOVSFWCLAAAGROILKIRKQFFHAIMRQEIOWFDVHVGELNRLTD 177
Qy 180 DYSKINEGIDKIGMFFHSIATFTFTGTVGTRGKMLTLVLAISPVGLSAAIWAKILS 239
Db 178 DVSQKINEGIDKIGMFFHSIATFTFTGTVGTRGKMLTLVLAISPVGLSAAIWAKILS 237
Qy 240 SFTDKELLAYAKAGAAVEVLAAITVIAFGQKKELRYNKNLEAKGIGIKKAITANI 299
Db 238 SFTDKELLAYAKAGAAVEVLAAITVIAFGQKKELRYNKNLEAKGIGIKKAITANI 297
Qy 300 SIGAFLLIYASALAFWGTGTVLSSEYSIGQVLTFFSVLIGAFSIGQASPSIEAFAN 359
Db 298 SIGAFLLIYASALAFWGTGTVLSSEYSIGQVLTFFSVLIGAFSIGQASPSIEAFAN 357
Qy 360 ARGAAEYIFKIIDNKPISIDYSKSGHKPDNIKNLEFNHVSFYSRKEVKILKGLNLKV 419
Db 358 ARGAAEYIFKIIDNKPISIDYSKSGHKPDNIKNLEFNHVSFYSRKEVKILKGLNLKV 417
Qy 420 QSGOTVALVGNCGGKSTTVQLMQRLYDPTGVMVSDGQDRTINVRHLREITGVVSOEP 479
Db 418 QSGOTVALVGNCGGKSTTVQLMQRLYDPTGVMVSDGQDRTINVRHLREITGVVSOEP 477
Qy 480 VLFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPNKFDTLVGERGQALSGGOK 539
Db 478 VLFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPNKFDTLVGERGQALSGGOK 537

QY 540 RIATARALVRNPKILLDEATSDALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADY 599
Db 538 RIATARALVRNPKILLDEATSDALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADY 597
QY 600 IAGFDDGVIVKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGSKSESDALEMSPKD 659
Db 598 IAGFDDGVIVKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGSKSESDALEMSND 657
QY 660 SGSLIKRRSTRSIIHAPQODRKLGTKEDLNENVPVPSFWRLLKLNSTWYFVVGIPFC 719
Db 658 SRSLIKRRSTRSRVGSQAQDRKLSKEALDESIPVPSFWRIMKLNLTWENYFVVGVC 717
QY 720 AIINGGLQAFSIFRSIIIGIFTRDEDPETKRONSNMFSVLFVLGLIISFIFTFLOQTF 779
Db 718 AIINGGLQAFSIFRSIIIGIFTRDEDPETKRONSNLFSLLFLALGISFIFTFLOQTF 777
QY 780 GKAGEILTKRLRYMFRSMLRQDVSWFDDPKNTGALTTRLANDAAQVKAGISRLAVIT 839
Db 778 GKAGEILTKRLRYMFRSMLRQDVSWFDDPKNTGALTTRLANDAAQVKAGISRLAVIT 837
QY 840 QNIANLGTGIIISLIYGWLTLLLLAIVPIIAIAGVVMKMLSGQALKKKELEGAGKIA 899
Db 838 QNIANLGTGIIISFIYGWLTLLLLAIVPIIAIAGVVMKMLSGQALKKKELEGAGKIA 897
QY 900 TEAIENFTVVSUTREQFEYMYAOSLOVPYRNSLRKAHIFGVFSFISITQAMMYFAGCF 959
Db 898 TEAIENFTVVSUTREQFEYMYAOSLOVPYRNSLRKAHIFGIFTSFTQAMMYFAGCF 957
QY 960 RFGAYLVANEFNFQDVLVFSIAVFGAMAVQVSPADYAKAVSAHVIMIEKSPL 1019
Db 958 RFGAYLVAKMLSFEDVLVFSIAVFGAMAVQVSPADYAKAKISAHHIIMIEKTPL 1017
QY 1020 IDSYSPHGLKPNLTGNVTFNFWFNYTPRDPDIPVLQGLSLEVKKGOTLALVSGSGCKS 1079
Db 1018 IDSYSEGLMPNTLEGNVTFGEVFNYPTRDPDIPVLQGLSLEVKKGOTLALVSGSGCKS 1077
QY 1080 TVVOLLERYDPLASVLIDGKEIKHLNVQWLRHLGLVISOEPILFDCSTAEINAYGDN 1139
Db 1078 TVVOLLERYDPLASVLIDGKEIKHLNVQWLRHLGLVISOEPILFDCSTAEINAYGDN 1137
QY 1140 RVVSHEEIMQAAKEANIHHFETLPEKYNTRVGDGKGTOLSGOKORTAIARALVRPHIL 1199
Db 1138 RVVSQEEIVRAKANIHAFTESLPNKYSTKVDGKGTOLSGOKORTAIARALVRPHIL 1197
QY 1200 LDEATSDALDTESEKVOEALDKAREGRTICIVIAHRLSTIONADLIYVFGKVKRGTH 1259
Db 1198 LDEATSDALDTESEKVOEALDKAREGRTICIVIAHRLSTIONADLIYVFGKVKRGTH 1257
QY 1260 QOLLAQKGIYFSMVSVQAGAKR 1281
Db 1258 QOLLAQKGIYFSMVSVQAGTKR 1279

RESULT 3
US-08-784-649A-2
; Sequence 2, Application US/08784649A
; Patent No. 5830697
; GENERAL INFORMATION:
; APPLICANT: Sikic, Branimir I
; APPLICANT: Chen, Gang
; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
; TITLE OF INVENTION: CYCLOSPORIN MODULATION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/784,649A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: Reg.No. 5830697 36,677
; REFERENCE/DOCKET NUMBER: 06037/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-322-5070
; TELEFAX: 415-854-0875
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1279 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-784-649A-2

Query Match 90.1%; Score 5835.5; DB 2; Length 1279;
Best Local Similarity 90.6%; Pred. No. 0;
Matches 1161; Conservative 55; Mismatches 61; Indels 5; Gaps 5;

QY 1 MDPEGKRGKSA-EKNFWKMGKKKKNEKKKPTVSTFAMFRYSNWLDRMLVGYTMAAI 59
Db 1 MDLEGDRNGAKKNFFKLNKSEK-DKKEKPTVSFEMFRYSNWLDRMLVGYTMAAI 59

QY 60 IHCAALPLMLVFGNMTDSFANAGISRNKTPPVIIINESITNTQHFINHLEEMTTVAY 119
Db 60 IHGAGLPLMLVFGMTDIFANAG-NLEDLMSNITNRSINDTGFEMN-LEEDMTRYAY 117

QY 120 YSGIAGVLVAAYIOVSFWCLAAAGROILKIRKQFPHALMROEIGFVHDVDELNRLTD 179
Db 118 YSGIAGVLVAAYIOVSFWCLAAAGROILKIRKQFPHALMROEIGFVHDVDELNRLTD 177

QY 180 DYSKINEGIDGKIGMFFHSIATFTFTGTVGTRGMKTLVILAIISPVLGSAAWAKILS 239
Db 178 DYSKINEVIGDKIGMFFOSMATFTFTGTVGTRGMKTLVILAIISPVLGSAAWAKILS 237

QY 240 SFTDKELIAYAKAGAAVEVLAAITVIAFGQKKELERYKNLEAKGIGIKKAITANI 299
Db 238 SFTDKELIAYAKAGAAVEVLAAITVIAFGQKKELERYKNLEAKGIGIKKAITANI 297

QY 300 SIGAAFLLIYASALAFWYGTSLVLSSEYSTIGQVLTVFVSFVLIGAFSIGOASPSIEAFN 359
Db 298 SIGAAFLLIYASALAFWYGTSLVLSSEYSTIGQVLTV-FSVLIGAFSIGOASPSIEAFN 356

QY 360 ARGAAEYEFKIIDNKPSIDSYSGKHGPDNITKGNLEFKNVHFYSYPSRKEYKILKGLNKKV 419
Db 357 ARGAAEYEFKIIDNKPSIDSYSGKHGPDNITKGNLEFKNVHFYSYPSRKEYKILKGLNKKV 416

QY 420 QSGQTVLVNCGSGCKSTTVOLMORLYDPTDGMVCIDGQDIRTINVRHLREITGVVSQEP 479
Db 417 QSGQTVLVNCGSGCKSTTVOLMORLYDPTDGMVSDGQDIRTINVRHLREITGVVSQEP 476

QY 480 VLFATTAEINIRYGRNVMTDEIKAYKEANAYDFIMKLPNKFDTLVGERGAOLSGGQK 539
Db 477 VLFATTAEINIRYGRNVMTDEIKAYKEANAYDFIMKLPNKFDTLVGERGAOLSGGQK 536

QY 540 RIATARALVRNPKILLDEATSDALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADY 599
Db 537 RIATARALVRNPKILLDEATSDALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADY 596

QY 600 IAGFDDGVIVKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGSKSESDALEMSPKD 659
Db 597 IAGFDDGVIVKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGSKSESDALEMSND 656

QY 660 SGSLIKRRSTRSIIHAPQODRKLGTKEDLNENVPVPSFWRLLKLNSTWYFVVGIPFC 719
Db 658 SGSLIKRRSTRSIIHAPQODRKLGTKEDLNENVPVPSFWRLLKLNSTWYFVVGIPFC 719

```

;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-9808
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-752-447-2

Query Match      89.9%; Score 5825; DB 2; Length 1280;
Best Local Similarity 90.1%; Pred. No. 0;
Matches 1155; Conservative 58; Mismatches 65; Indels 4; Gaps 4;

QY 1 MDPEGGRGKSA-EKNFWKMGKSKKNEKKKPTVSTFAMFRYSNWLDRILMLVGTMAAI 59
DB 1 MDLEGDRNGAKKNFKLNKSEK-DKKEKKPTVSVFSEFRYSNWLDRILMLVGTMAAI 59

QY 60 IHGAALPLMLVFGNMTDSFANAGISRNKTFPPVIINSEINNTQHFNIHLEEMTYYAY 119
DB 60 IHGAGLPLMLVFGEMTDIFANAG-NLEDLSNITNRSRINDTGFNMN-LEEDMTRYYAY 117

QY 120 YSGTGAGVLVAAAYTOVSFWCLAAAGROILKIRKQFFHAIMRQEIQWFDVHDVDELNRLTD 179
DB 118 YSGTGAGVLVAAAYTOVSFWCLAAAGROILKIRKQFFHAIMRQEIQWFDVHDVDELNRLTD 177

QY 180 DVSKINEGIDKIGMFHSHSIATFTFTGTVGTRGKWLTVLILASPVILGSLAAIWKILS 239
DB 178 DVSKINEGIDKIGMFQSMATFTFTGTVGTRGKWLTVLILASPVILGSLAAIWKILS 237

QY 240 SFTDKELLAYAKAGAAVEEVLAARITVIAFGGOKKELRYNKNLEAKGIGIKKAITANI 299
DB 238 SFTDKELLAYAKAGAAVEEVLAARITVIAFGGOKKELRYNKNLEAKGIGIKKAITANI 297

QY 300 SIGAFLLLIYASYALAFWGTSLVLSSEYSIGQVLTFFSVLIGAFSGQASPIEAFAN 359
DB 298 SIGAFLLLIYASYAVAFWGTSLVLSSEYSIGQVLTFFSVLIGAFSGQASPIEAFAN 357

QY 360 ARGAAYEIFKIIDNKPSIDSYSGSKGHKPDNKGNEFNHVSFYSRKEVKILGLNLKV 419
DB 358 ARGAAYEIFKIIDNKPSIDSYSGSKGHKPDNKGNEFNHVSFYSRKEVKILGLNLKV 417

QY 420 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDRTINVRHLREITGVVSQEP 479
DB 418 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVSDGQDRTINVRHLREITGVVSQEP 477

QY 480 VLFATTIAENIRYGRNVTMDIEIKAVKANAYDFIMKLPNKFDTLVGERGAQLSGGQK 539
DB 478 VLFATTIAENIRYGRNVTMDIEIKAVKANAYDFIMKLPNKFDTLVGERGAQLSGGQK 537

QY 540 RTAIAALVNRNPKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 599
DB 538 RTAIAALVNRNPKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 597

QY 600 IAGFDGVTIVEKGNHDELMEKRGIFPKVTMTGRNEIELENATGESKESDALEMSPKD 659
DB 598 IAGFDGVTIVEKGNHDELMEKRGIFPKVTMTGRNEIELENATGESKESDALEMSND 657

QY 660 SGSSLIKRSTRSRSHIAPQOQDKRLGTKEADLENVPPVSFWRLTKLNSTWEPYFVVGIFC 719
DB 658 SRSSLIKRSTRSRVSGSQAHKRLSTKEALDESIPVSVFWRLTKLNSTWEPYFVVGIFC 717

QY 720 AINGGLOPAFSTIIFSRIGIFTRDEDPETKRONSNMFSVLFLVLGIISFTIFFLOGFTF 779
DB 718 AINGGLOPAFAIFSKIIIGVFRIDDPETKRONSNLFSLLFLALGIISFTIFFLOGFTF 777

QY 780 GKAGEILTRELRYMFRSMRLRQDVSFDDPKNTTGALTFLANDAAQVGAIGSLRAVIT 839
DB 778 GKAGEILTRELRYMFRSMRLRQDVSFDDPKNTTGALTFLANDAAQVGAIGSLRAVIT 837

QY 840 QNTANLGTGIIISLIYQWLTLLALLVPIIAIAGVVEKMLSGQALKDKKELEGAGKIA 899
;
; REFERENCE/DOCKET NUMBER: 95,1121

```

```

;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-9808
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-752-447-2

Query Match      89.9%; Score 5825; DB 2; Length 1280;
Best Local Similarity 90.1%; Pred. No. 0;
Matches 1155; Conservative 58; Mismatches 65; Indels 4; Gaps 4;

QY 1 MDPEGGRGKSA-EKNFWKMGKSKKNEKKKPTVSTFAMFRYSNWLDRILMLVGTMAAI 59
DB 1 MDLEGDRNGAKKNFKLNKSEK-DKKEKKPTVSVFSEFRYSNWLDRILMLVGTMAAI 59

QY 60 IHGAALPLMLVFGNMTDSFANAGISRNKTFPPVIINSEINNTQHFNIHLEEMTYYAY 119
DB 60 IHGAGLPLMLVFGEMTDIFANAG-NLEDLSNITNRSRINDTGFNMN-LEEDMTRYYAY 117

QY 120 YSGTGAGVLVAAAYTOVSFWCLAAAGROILKIRKQFFHAIMRQEIQWFDVHDVDELNRLTD 179
DB 118 YSGTGAGVLVAAAYTOVSFWCLAAAGROILKIRKQFFHAIMRQEIQWFDVHDVDELNRLTD 177

QY 180 DVSKINEGIDKIGMFHSHSIATFTFTGTVGTRGKWLTVLILASPVILGSLAAIWKILS 239
DB 178 DVSKINEGIDKIGMFQSMATFTFTGTVGTRGKWLTVLILASPVILGSLAAIWKILS 237

QY 240 SFTDKELLAYAKAGAAVEEVLAARITVIAFGGOKKELRYNKNLEAKGIGIKKAITANI 299
DB 238 SFTDKELLAYAKAGAAVEEVLAARITVIAFGGOKKELRYNKNLEAKGIGIKKAITANI 297

QY 300 SIGAFLLLIYASYALAFWGTSLVLSSEYSIGQVLTFFSVLIGAFSGQASPIEAFAN 359
DB 298 SIGAFLLLIYASYAVAFWGTSLVLSSEYSIGQVLTFFSVLIGAFSGQASPIEAFAN 357

QY 360 ARGAAYEIFKIIDNKPSIDSYSGSKGHKPDNKGNEFNHVSFYSRKEVKILGLNLKV 419
DB 358 ARGAAYEIFKIIDNKPSIDSYSGSKGHKPDNKGNEFNHVSFYSRKEVKILGLNLKV 417

QY 420 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDRTINVRHLREITGVVSQEP 479
DB 418 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVSDGQDRTINVRHLREITGVVSQEP 477

QY 480 VLFATTIAENIRYGRNVTMDIEIKAVKANAYDFIMKLPNKFDTLVGERGAQLSGGQK 539
DB 478 VLFATTIAENIRYGRNVTMDIEIKAVKANAYDFIMKLPNKFDTLVGERGAQLSGGQK 537

QY 540 RTAIAALVNRNPKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 599
DB 538 RTAIAALVNRNPKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 597

QY 600 IAGFDGVTIVEKGNHDELMEKRGIFPKVTMTGRNEIELENATGESKESDALEMSPKD 659
DB 598 IAGFDGVTIVEKGNHDELMEKRGIFPKVTMTGRNEIELENATGESKESDALEMSND 657

QY 660 SGSSLIKRSTRSRSHIAPQOQDKRLGTKEADLENVPPVSFWRLTKLNSTWEPYFVVGIFC 719
DB 658 SRSSLIKRSTRSRVSGSQAHKRLSTKEALDESIPVSVFWRLTKLNSTWEPYFVVGIFC 717

QY 720 AINGGLOPAFSTIIFSRIGIFTRDEDPETKRONSNMFSVLFLVLGIISFTIFFLOGFTF 779
DB 718 AINGGLOPAFAIFSKIIIGVFRIDDPETKRONSNLFSLLFLALGIISFTIFFLOGFTF 777

QY 780 GKAGEILTRELRYMFRSMRLRQDVSFDDPKNTTGALTFLANDAAQVGAIGSLRAVIT 839
DB 778 GKAGEILTRELRYMFRSMRLRQDVSFDDPKNTTGALTFLANDAAQVGAIGSLRAVIT 837

QY 840 QNTANLGTGIIISLIYQWLTLLALLVPIIAIAGVVEKMLSGQALKDKKELEGAGKIA 899
;
; REFERENCE/DOCKET NUMBER: 95,1121

```

```

;
; REFERENCE/DOCKET NUMBER: 95,1121

```

Db 838 QNANLGTGIIISFYGNQLTLLLLAIVPIIAIAGVVMKMFAGQALKDKKELEGAGKIA 897
Qy 900 TEATENFTVYSLTREKFEYMAQSLQVPYRNSLRKAHIFGVYSFSTQAMMYFSYAGCF 959
Db 898 TEATENFTVYSLTREKFEYMAQSLQVPYRNSLRKAHIFGVYSFSTQAMMYFSYAGCF 957
Qy 960 RFGAYLVANEFMNPQDVLVFSIAIVFGAMAVGOVSSFPADYAKAKVSAAHVIMIEKSPL 1019
Db 958 RFGAYLVANEFMNPQDVLVFSIAIVFGAMAVGOVSSFPADYAKAKVSAAHVIMIEKSPL 1017
Qy 1020 IDSVSPGLKPNLTLEGNTVNEVFNYPTRPDIPVLOGLSLEKVKGGTALVSSGCGKS 1079
Db 1018 IDSVSPGLKPNLTLEGNTVNEVFNYPTRPDIPVLOGLSLEKVKGGTALVSSGCGKS 1077
Qy 1080 TVQOLLERFYDPLAGSLVLDGKEIKHLNVQWLAHGLGIVSQEPILPDCSTAEINAYGDNS 1139
Db 1078 TVQOLLERFYDPLAGSLVLDGKEIKHLNVQWLAHGLGIVSQEPILPDCSTAEINAYGDNS 1137
Qy 1140 RVYSHEIMQAKEANTHHFETLPEKYNTRVGDGKGTQLSGGQKQRIATARALVROPHIL 1199
Db 1138 RVYSHEIMQAKEANTHHFETLPEKYNTRVGDGKGTQLSGGQKQRIATARALVROPHIL 1197
Qy 1200 LLEATSALETSKVVQVQALDKAREGTCIVIAHRLSTIQNADLIWVFQNGVKVKEHGT 1259
Db 1198 LLEATSALETSKVVQVQALDKAREGTCIVIAHRLSTIQNADLIWVFQNGVKVKEHGT 1257
Qy 1260 QQLAQKGIYFSMVSVQAGAKR 1281
Db 1258 QQLAQKGIYFSMVSVQAGTKR 1279

RESULT 5

US-09-120-513-2
; Sequence 2, Application US/09120513
; Patent No. 6025160
; GENERAL INFORMATION:
; APPLICANT: Brun, Kimberly
; APPLICANT: Chenery, Richard
; APPLICANT: Ellens, Harma
; APPLICANT: Field, John
; APPLICANT: Yue, Lin
; TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE
; TITLE OF INVENTION: SEQUENCES ENCODING RAT MDR1B2 AND
; TITLE OF INVENTION: SCREENING METHODS THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: US
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120.513
; FILING DATE: 22-JUL-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: King, William T
; REGISTRATION NUMBER: 30,954
; REFERENCE/DOCKET NUMBER: GP50008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5015
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1275 amino acids
; TYPE: amino acid

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-120-513-2

Query Match 81.6%; Score 5286.5; DB 3; Length 1275;
Best Local Similarity 80.3%; Pred. No. 0;
Matches 1029; Conservative 130; Mismatches 114; Indels 9; Gaps 5;

Qy 1 MPEGGKGSABKNFMWKKSKKNEKKKPTVSTFAMFRYSNNWLDRLYLVLGVTMAAIL 60
Db 1 MEFEELNGTADKNFSKMGKSKK--EKKKPAVGIFGMFRYADWLDKLCMALGTLAAIL 58
Qy 61 HGAALPLMLVGNMTDSFANAGISRKNKTPFVINESTNTQHF--NHLEEMTYAYY 119
Db 59 HGTLPPLMLVGYMTDSFTQA---ETRLPVTNQSEINSTQTVSDSLEEDMAYAYY 115
Qy 120 YSGIAGVLVAAYIQVSWCLAAAGROIKIRKOFFHAIMROEIGWDFDHDVDELNTRLTD 179
Db 116 YTGIGAGVLVAYIQVSWCLAAAGROIKIRKOFFHAIMROEIGWDFDHDVDELNTRLTD 175
Qy 180 DYSKINEGIDGKIGMFFHSIAFTFTGIVGFTGRGKLTVLIAISPLVLGSAAIWAKILS 239
Db 176 DYSKINEGIDGKIGMFFHSIAFTFTGIVGFTGRGKLTVLIAISPLVLGSAAIWAKILS 235
Qy 240 SFTDKELLAYAKAGAVAEVLAARIPTVAFGQKKELERYNKNLEEAIGIKKAITANI 299
Db 236 SFTDKELLAYAKAGAVAEVLAARIPTVAFGQKKELERYNKNLEEAIGIKKAITANI 295
Qy 300 SIGAFLLIYASYALAFWYGTSLVLSSEYSIGOVLTVPFVSLVIGAFSIGQASPSPEAFN 359
Db 296 SIGAFLLIYASYALAFWYGTSLVLSSEYSIGOVLTVPFVSLVIGAFSIGQASPSPEAFN 355
Qy 360 ARGAAEYIFKIIDNKPISDSYSGHKKPDNKGLEKKNVHFSYPSRKEVYKILGNLKV 419
Db 356 ARGAAEYIFKIIDNKPISDSYSGHKKPDNKGLEKKNVHFSYPSRKEVYKILGNLKV 415
Qy 420 OSGQTVLVGNSGCGKSTTVQLMQRLYDPTGDMVCIDQDRTINVRHLREITGVVSOEP 479
Db 416 OSGQTVLVGNSGCGKSTTVQLMQRLYDPTGDMVCIDQDRTINVRHLREITGVVSOEP 475
Qy 480 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQK 539
Db 476 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQK 535
Qy 540 RIATARALVRNPKILLDEATSALETSKVVQVQALDKAREGTCIVIAHRLSTVRNADY 599
Db 536 RIATARALVRNPKILLDEATSALETSKVVQVQALDKAREGTCIVIAHRLSTVRNADY 595
Qy 600 IAGFDDGVIVVEKGNHDELMKEGIVFKLVMTQTRNGNELELENATGESKESDALEMSPKD 659
Db 596 IAGFDDGVIVVEKGNHDELMKEGIVFKLVMTQTRNGNELELENATGESKESDALEMSPKD 655
Qy 660 SGSLIKRRSTRRSIHAPQGGQDKRLGKEDLNENVPVPSFWIRLKLNSTEWPYVVGIFC 719
Db 656 SKSPLI--RRSIRRSIHRRQDQERLSKEDVEDVPMVFWOILKLNISEWPIYVGVLC 714
Qy 720 AINGGLQPAFSIIFSRIGIIFTRDEDEPEYKQNSMFSVLFLVLGIISFITFFLQGFTE 779
Db 715 AVINGCIQPVFAIVFSKIVGVFSRDDHETKQRNCNLFSLFLVMGISEVTVFFQGFTE 774
Qy 780 GKAGEILTKRLRYMVFRSMLRQDVSFDDPKNTTGALTITRLANDAQVKAIGSRRAVIT 839
Db 775 GKAGEILTKRLRYMVFRSMLRQDVSFDDPKNTTGALTITRLANDAQVKAIGSRRAVIT 834
Qy 840 QNANLGTGIIISLYGWQLTLLLLAIVPIIAIAGVVMKMFAGQALKDKKELEGAGKIA 899
Db 835 QNANLGTGIIISLYGWQLTLLLLAIVPIIAIAGVVMKMFAGQALKDKKELEGAGKIA 894
Qy 900 TEATENFTVYSLTREKFEYMAQSLQVPYRNSLRKAHIFGVYSFSTQAMMYFSYAGCF 959
Db 895 TEATENFTVYSLTREKFEYMAQSLQVPYRNSLRKAHIFGVYSFSTQAMMYFSYAGCF 954

QY 960 RFGAYLVANEFNFODVLLVSAIVFGAMAYGOVSSFPADYAKAKVSAHVIMLIEKSPL 1019
D 955 RFGAYLVARELMTFENVMLVSAVFGAMAAAGNTSSFPADYAKAKVSAHVIMLIEKSPL 1014
QY 1020 IDSYSPHGLKPNLTGNTVFNVEVFNYPTRPDIPVLOGLSLEVKGQTLALVSGSGCGKS 1079
D 1015 IDSYSTEGGLKPNWLGNGVFNKFGVFNYPTRPNIPVLOGLSLEVKGQTLALVSGSGCGKS 1074
QY 1080 TVVQLLERYDPPLAGSVLDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAENTAYGNS 1139
D 1075 TVVQLLERYPNMAGTVFLDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAENTAYGNS 1134
QY 1140 RVVSHETIMQAQKANIHFETLPEKYNTRVGDGKTOLSGGOKORIAIALRVQPHIL 1199
D 1135 RVVSHETIVRAAREANIHQFIDSLPEKYNTRVGDGKTOLSGGOKORIAIALRVQPHIL 1194
QY 1200 LLDATSALDTESEKVVQVQALDKAREGRTCVIAHRLSTIQNADLIVVQNGVKEHGH 1259
D 1195 LLDATSALDTESEKVVQVQALDKAREGRTCVIAHRLSTIQNADLIVVQNGVKEHGH 1254
QY 1260 QQLLAQKGIYFSMVQVQAKR 1281
D 1255 QQLLAQKGIYFSM--VQAKR 1274
RESULT 6
US-09-450-105-2
; Sequence 2, Application US/09450105
; Patent No. 6169166
; GENERAL INFORMATION:
; APPLICANT: Kimberly Anne Brun
; APPLICANT: Richard James Chenery
; APPLICANT: Harma Ellens
; APPLICANT: John Anthony Feild
; TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE SEQUENCES
; FILE OF INVENTION: ENCODING RAT MDR1B2 AND SCREENING METHODS THEREOF
; FILE REFERENCE: GP-50008-D1
; CURRENT APPLICATION NUMBER: US/09/450,105
; CURRENT FILING DATE: 1999-11-29
; EARLIER APPLICATION NUMBER: 09/120,513
; EARLIER FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1275
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-450-105-2
Query Match 81.6%; Score 5286.5; DB 4; Length 1275;
Best Local Similarity 80.3%; Pred. No. 0;
Matches 1029; Conservative 130; Mismatches 114; Indels 9; Gaps 5;
QY 1 MDPBGRKGSAGNFWKMGKSKKNEKKKPTVSTFAMFRYNWLDRLVLMVGTMAAII 60
D 1 MEFPEGLNGTADNFSKMGKSKK--EKEKPAVGIFGFRYADWLDKLCMALGTAAII 58
QY 61 HGAALPLMLVFGNMTDSFANAGISRKNKTPFVIINSEITNTQHTI-NHLEEEEMTYAYY 119
D 59 HGTLLPLMLVFGNMTDSFQA---ETRIILPVTNNGSEINSTQVSDSSLEEDMAMAYY 115
QY 120 YSGIGAGVLVAAAYIQVSWFLAAGRAQILKIRKQFFHAIMRQEIWFVDVHDVGNELRLTD 179
D 116 YTGIGAGVLVAYIQVSWFLAAGRAQILKIRKQFFHAIMRQEIWFVDVHDVGNELRLTD 175
QY 180 DVSINIGIDGKIGMFPHSSTATFTGIVGTRGKWLTVLILASPVLGLSAAIWAIIKS 239
D 176 DVSINIGIDGKIGMFPHSSTATFTGIVGTRGKWLTVLILASPVLGLSAAIWAIIKS 235
QY 240 SFTDKELLAYAKAGAAVEELAAIRTVIAFGQKKELERYKNLEAKGIGIKKAITANI 299
D 236 SFTDKELLAYAKAGAAVEELAAIRTVIAFGQKKELERYKNLEAKGIGIKKAITANI 295

QY 300 SIGAAFLLIYASVALAFWYGTSLVLSSEYSIGQVLTVFFSVLIGAFSICQASPSIEAFAN 359
D 296 SIGIAYLLVYASVALAFWYGTSLVLSSEYSIGQVLTVFFSVLIGAFSICQASPSIEAFAN 355
QY 360 ARGAAVYFIKIDNKPSIDYSKSGHKPNKGNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKN 419
D 356 ARGAAVYFIKIDNKPSIDYSKSGHKPNKGNLKNLKNLKNLKNLKNLKNLKNLKNLKNLKN 415
QY 420 QSGOTVALVNSGCGKSTTVQLMORLYDPTDMVCIDGQDIRTINVRHLREITGVVSOEP 479
D 416 KSGOTVALVNSGCGKSTTVQLMORLYDPTDMVCIDGQDIRTINVRHLREITGVVSOEP 475
QY 480 VLFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPNKEDTLVGERCAQLSGGQK 539
D 476 VLFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPNKEDTLVGERCAQLSGGQK 535
QY 540 RIATARALVRNPKLILLDEATSAIDTESEAVVOALDKARKGRTTIVIAHRLSTVNRADV 599
D 536 RIATARALVRNPKLILLDEATSAIDTESEAVVOALDKARKGRTTIVIAHRLSTVNRADV 595
QY 600 IAGFDGVIYKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGESKESDSEDALESPKD 659
D 596 IAGFDGVIYKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGESKESDSEDALESPKD 655
QY 660 SGSSLIKRRSTRSIHAPQGDRLKTKEDLNENVPVSWRILKLNSTWMPYFVVCIFC 719
D 656 SKSPLI--RRSIRRSIHRROQERRLSKEDVEDVPMVSFWQILKLNISEMPLVVGVL 714
QY 720 AINGGLQPAFISIIIGITRDEDEPKTRNSNMFVFLVGLVIGIISITFTFLOQFTF 779
D 715 AVINGCIPQVFAIVFSKIVGVSRDDHDEYKORNCNLFSLFLVGMISVFTVFFQGT 774
QY 780 KGAGEILTKRLRYMVRSMRLQDVSDFDDPKNTTGALTURLANDAAOVKGAISRLAVIT 839
D 775 KGAGEILTKRLRYMVRSMRLQDVSDFDDPKNTTGALTURLANDAAOVKGAISRLAVIT 834
QY 840 QNIANLGTGIIISLYGWQLTLLLAIPITAIAGVEMKMLSGOALKDKKELEGACKIA 899
D 835 QNVANLGTGIIISLYGWQLTLLLAIPITAIAGVEMKMLSGOALKDKKELEGACKIA 894
QY 900 TEATENFTVSLTREQFEYMYAQSLOVYPYRNSLRKAHIFGVSTFQAMMYFSYAGCF 959
D 895 TEATENFTVSLTREQFEYMYAQSLOVYPYRNSLRKAHIFGVSTFQAMMYFSYAGCF 954
QY 960 RFGAYLVANEFNFODVLLVSAIVFGAMAYGOVSSFPADYAKAKVSAHVIMLIEKSPL 1019
D 955 RFGAYLVARELMTFENVMLVSAVFGAMAAAGNTSSFPADYAKAKVSAHVIMLIEKSPL 1014
QY 1020 IDSYSPHGLKPNLTGNTVFNVEVFNYPTRPDIPVLOGLSLEVKGQTLALVSGSGCGKS 1079
D 1015 IDSYSTEGGLKPNWLGNGVFNKFGVFNYPTRPNIPVLOGLSLEVKGQTLALVSGSGCGKS 1074
QY 1080 TVVQLLERYDPPLAGSVLDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAENTAYGNS 1139
D 1075 TVVQLLERYPNMAGTVFLDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAENTAYGNS 1134
QY 1140 RVVSHETIMQAQKANIHFETLPEKYNTRVGDGKTOLSGGOKORIAIALRVQPHIL 1199
D 1135 RVVSHETIVRAAREANIHQFIDSLPEKYNTRVGDGKTOLSGGOKORIAIALRVQPHIL 1194
QY 1200 LLDATSALDTESEKVVQVQALDKAREGRTCVIAHRLSTIQNADLIVVQNGVKEHGH 1259
D 1195 LLDATSALDTESEKVVQVQALDKAREGRTCVIAHRLSTIQNADLIVVQNGVKEHGH 1254
QY 1260 QQLLAQKGIYFSMVQVQAKR 1281
D 1255 QQLLAQKGIYFSM--VQAKR 1274

RESULT 7
US-08-612-734B-2
; Sequence 2, Application US/08612734B

```

: Patent No. 5914246
: GENERAL INFORMATION:
: APPLICANT: Peery, Robert B.
: APPLICANT: Skatrud, Paul L.
: APPLICANT: Tobin, Matthew B.
: TITLE OF INVENTION: Multiple Drug Resistance Gene of
: TITLE OF INVENTION: Aspergillus Fumigatus
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Eli Lilly and Company
: STREET: Lilly Corporate Center, DC1501
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: U.S.A.
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/612,734B
: FILING DATE: 08-MAR-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Craig, Anne I.
: REGISTRATION NUMBER: 32,976
: REFERENCE/DOCKET NUMBER: X-9681
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-354-9570
: TELEFAX: 617-354-4043
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1349 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-612-734B-2

Query Match      38.8%; Score 2511; DB 2; Length 1349;
Best Local Similarity 42.5%; Pred. No. 4e-228;
Matches 563; Conservative 231; Mismatches 458; Indels 74; Gaps

QY      1  MDPGGKK---GSA-EKNFWKMGKKKKNNKKKK-----PT--VSTFAMFRYSNNWLDRL 49
Db      1  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      50  MAPDGGKKKHDKGKAVDNDLSLFAHLQEHKEVYLRQLDAPSVKVSFFTYLYRASKDIL 109
Db      50  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      50  YMLVGTMAALIHGAALPLMLFGNMTDSPANAGIRSNKTFPVIINESITNTQHEINHL 109
Db      50  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      110  IILVSAICATAAGAAALPLFILFGSLASAF--QGSLG-TMP-----YHEFYH- 154
Db      110  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      110  EEEMTYAYYSGIGAGVLVAAYIQVSWFCLAGROILKIRKOFFHAIMRQEIFGWFDVDH 169
Db      110  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      155  --LKNLVLFVYVYLGTAEFVTVVSTVGFYITGHEHTQKIRENYLEAILRQNNMAYFDKLG 212
Db      155  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      170  VGEINLRLTDVSKINEGIDKIGMFFHSTATFTGTFIVGFRGWKLTIV----ILAISP 225
Db      170  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      213  AGEVTFRITADTNLIQDAISEKVGCLTLTAFATVFAIVAYVYKWLALCTICTSTIVALVM 272
Db      213  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      226  VLGLSRAIWAKLSSPTFDKELIAYAKAGAAVEVLAAIRTVIAFGQKKELERYKNKLEE 285
Db      226  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      273  VMGGG----SRFIVKYSKSISSYAGGTVAAEVISSIRNATAFGTQDKLAKQYETHLAE 328
Db      273  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      286  AKGIGIKKAITANISGAAPFLIYASALAFWYGTSLVLSSEYSQGVLTVFVSVLIGAF 345
Db      286  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      329  AEKWGVKQVVLGMIMTGMFGFMFNGYLGFWMGSRFVVGKENVYGOVLTVLMLSLGFSF 388
Db      329  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      346  SIGQASPSTEAFANARGAAEYFIKIDNKPSSIDTSYKSGHKPDNTKGNLEKFNHVSYP 405
Db      346  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      389  SLGNVAPNGQAFTNGVAAAAKIYSTIDRSPLDPTSDGKVLDFHPEGNIERNKHVHTPS 448
Db      389  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      406  RKEVKILKGLNLKVGSGQTVALVGNSSGCKSTTVOLMQRLLYDPTDGMVICDQDITINV 465
Db      406  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db	449	RPEVTVMEDVSI SMPAGKTKTALVGGSGGKSTTVVGLVERFYLPVGGQVLDDGHDIOTLNL	508
QY	466	RHLREITGVVSEPPVLFATTAENIRYGR-----ENVTMDEI-----EKAVKEANAYDFIM	516
Db	509	RWLROQISLVVSEPPVLFSTTIFRNIHGLIGTKFEHESKOKIRELVENARMANAHDFIM	568
QY	517	KLPNQFDTLVGERGAQLSGGQKQRIATARALVRNPKLILLDEATSAIDTSEAVVQVALD	576
Db	569	ALUEGYDTNWGQGFLLSGGQKQRIATARATVSDPKLILLDEATSAIDTSEGVQQAALD	628
QY	577	KARKGRTTIVIAHRLSTVRNADVITAGFDGQVIVIEKGHNHLMKEKGYFKFLYKTMOTRGNE	636
Db	629	KAAREGRTTIVIAHRLSTIKTAHNIVAMVGGKIAOQTHDELVDKRGTYKYLVEAQINSEE	688
QY	637	IELENATGESKESDALEMSPKDSSGLIKR-----RSTRSIHAPOQDQR-----KLGTKE	688
Db	689	KEAE-----ALEADA-DMDADDFEQEGVTRIKTAVSSSSNLSDAVDEKARLEMKRGTQK	741
QY	689	D-----LNENVP-----PVSEFWRLK-----LNSTPEWPFVVGIFCAILINGLOPAISIFPS	735
Db	742	SVSSAVLSKKVPEQFEKYSLMTIVKFCANRPRLGMLIGLFFSLAGGGQPTQAFIYA	801
QY	736	RIIGFTRDEDPETK-RONSNMFSVLFLVLGIISITFTFFLQGGFTFGKAGEIITLKLRYMV	794
Db	802	KAISTLSLPSMFHKLHRDANFWSLMPFVVGIAQFISLSINGTAPAFICSERLIRRASQA	861
QY	795	FRSMLRQDVSFNDPKNTGALTITRLANDAAQVKGAGSLAVITONIANLGTGIIISLI	854
Db	862	FRSILURQDISFDDREENTGALTSFSLSTETKNLSGVSGVTLGTITMTSTLGAAMIATA	921
QY	855	YGWQLTLLLLLAIVPIIAIAGVVEKMLSGOALKDKKELEGAKIAPEATENRTFVVSLTR	914
Db	922	IGWKLALVCISVVPILLACGFLREYMLAQFOQRSKSAYEGSASYACEATSAIRTVASLTR	981
QY	915	EOKFEYNTAQSLQVPPYRNSLRKAHIFGVSPSITQAMMYFSYACGFRFGAYLVANEFMNFQ	974
Db	982	EQDVWGVYHDLQKQGRKSLISVLRSSLLYASSQALVFFCVAGLGFYWGGLTGLGHEYSIF	1041
QY	975	DVLLVFSIAIVFGAMAGOVSSFAPDYAKAKVSAAHVIMIEKSPILDYSVSPHLKPNTLE	1034
Db	1042	RFFVCFSEILFGASQAGTVSFAPDMCKAKANAAQFKFLDSKPTIDLWSDGEKLESM	1101
QY	1035	GNVTFNVFVFNPRDPIDPVLQGLSLEVKKGQTLALVGGSGGKSTTVOLLERFYDPLAG	1094
Db	1102	GEIEFRDWHFYPTRPQPVLRGLNLSVKPGQYIALVGPSCGKSTTALLERFYDALAG	1161
QY	1095	SVLDGKEIKHLNVOWLRHGLTVSQEPIILFDCSTAEINTAYGDSNRVVSHEEIMQAKEA	1154
Db	1162	GVFDGDKIYKLVNVSYSRFLSIVSQEPHLYOGTITKENILLGVGDKDDVSEETLTKVCKDA	1221
QY	1155	NIHFITLPEKYNVRGDKGTOLSGGQKQRIARALVRQPHILLDEATSAIDTSEK	1214
Db	1222	NIYDFVMSLPEGFDTVVGGSGKMLSGQKQRIARALLRDPKVLLEDATSAIDTSEK	1281
QY	1215	VQOALDKAREGRTCTIVIAHRLSTIONADLIVVFQNGKVEHGTHQOOLAKGIFYFWSVS	1274
Db	1282	VQOALDAAARGRTTIAVAHRUSTIONADIIVFDQGIKESGTHHELIRNKRYELVN	1341
QY	1275	VQAGAK 1280	
Db	1342	LQSLGK 1347	
RESULT 8			
US-08-996-545-2			
; Sequence 2, Application US/08996545			
; Patent No. 5928898			
; GENERAL INFORMATION:			
; APPLICANT: Skatrud, Paul L.			
; APPLICANT: de Waard, Maarten A.			
; APPLICANT: Peery, Robert B.			
; APPLICANT: Andrade, Alan C.			
; TITLE OF INVENTION: Multiple Drug Resistance Gene atnd of			

```

RESULT 8
US-08-996-545-2
; Sequence 2, Application US/08996545
; Patent No. 5928898
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Dr

```

;; TITLE OF INVENTION: Aspergillus nidulans
;; NUMBER OF SEQUENCES: 3
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Eli Lilly and Company
;; STREET: Lilly Corporate Center
;; CITY: Indianapolis
;; STATE: Indiana
;; COUNTRY: U.S.
;; ZIP: 46285
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/996,545
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Webster, Thomas D.
;; REGISTRATION NUMBER: 39,872
;; REFERENCE/DOCKET NUMBER: X-11766
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 317-276-3334
;; TELEFAX: 317-276-2763
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1334 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-996-545-2

Query Match 38.3%; Score 2481; DB 2; Length 1334;
Best Local Similarity 42.0%; Pred. No. 2.7e-225;
Matches 538; Conservative 223; Mismatches 456; Indels 62; Gaps 8;

QY 24 KNEKKEKPVSTFAMFRYSNWLDRMLVGLTMAAIIHGAAALPMLMFGNMTDSFANAG 83
DB 90 KTOLEEKVNIISFFGLWRATKMDILIMVISTICAIAAASTFORIMLY----- 137

QY 84 ISRNKTFPVIINESNTNQHFTHLHEEMTTVAYYSIGAGVLVAAAYIOVSFWCLAAG 143
DB 138 -----QISYDEFYDELTKNVLVYVLGIGEFYTVVSTVGVFYTGE 178

QY 144 RQILKIRKQFFHALMROEIGMEDVDHVDGELNTRLTDDVSKINEGIDGKIGMFFHSTATFF 203
DB 179 HATQKIREYLESILRONIGYFDKLGAGEVTTTRITADTNLIQDGISEKVGLTLTALATFV 238

QY 204 TGFTVGTGRGKLTVLTAISPVLGSAALWAKILSSFTDKELLAYAKAGAAVEVLAAL 263
DB 239 TAFIIAVYKWKALICSSITIVLVTMGGSQFIKYSKSLDSYGAGGTVAEEVISSI 298

QY 264 RTVIAFGGKKELEERYKNLEAKGIGIKKAITANTISIGNAFLIIYASALAFWYGTSLV 323
DB 299 RNATAFGTQDLAKOYEVHLEDEAKETKQIVMGFMGAMFGLMYSNYGLGFMWMSRFL 358

QY 324 LSSEYSIGVLTVPFSLIGAFSIGOASPIEAFANARGAAEYIFKIIDNKPSIDSYSKS 383
DB 359 VDGAVDGDILTVLMAILIGSFLGNSVPAQFTNVAARAAKIFGTIDRQSPLDPSYNE 418

QY 384 GHKPDNKGKLEFNHFSYPSRKEVKILKGLNLKVOSGGTVALVNGSGCGKSTTVOLMQ 443
DB 419 GKTLDHFEGHIELRNKVIHYPSEVTVMEDVSLMPAGKTALVGPSPGSKSTVVGGLVE 478

QY 444 RLYDPTDGMVDCDQDRTINVRHLREITGVVSOEPLVLEATTAEINRYG-----RENV 498
DB 479 RFYMPVRGTVLVDGHDIKDLMLRWLRQOISLVSOEPLVFGTITVYKRNHGLIGTKYENES 538

QY 499 MDE-----IEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGQKQRTAIARALVRNPKIL 554
DB 539 EDKVRLELIENAKMANAHDFTALPEGYETNVGQGFLLSGQKQRTAIARAVVSDPKIL 598

QY 555 LLDDEATSDALTESEAVVQVALDKARKGRRTIVIAHRLSTVRNADVIAGFDDGVIVKEGNH 614
DB 599 LLDDEATSDALTESEAVVQVALDKARKGRRTIVIAHRLSTVRNADVIAGFDDGVIVKEGNH 658
QY 615 DELMKEKGIYFKLVMTQTRGNEIE---LENATGESKESDALEMSPKDSSGSS-----L 664
DB 659 DELVDRGGAYRKLVKVEAQRIEQAEDAEADAEADLTNADIAKIKTASSASSDLDCGKPTT 718
QY 665 IKRRSTRRSIHAPQGGQDRKLGTKEDLNENVPVPSFWRLK----LASTEPPYFVVGIFCA 720
DB 719 IDRTGTHKSVSSAILSKRP-----PETPKYSLMTLLKFVASFNRPEIPYMLLGLVFS 771
QY 721 IINGLOPAPSIIFSRIGITITRDEDEPETK-RONSNMFWSLFLVLGLISITITFFLOGFTF 779
DB 772 VLAGGQPTQAVLYAKAISTLSLPESQYSLKRDHDAFWSLFFVLIQITOSTNGAAF 831
QY 780 GKAGEILTCLRVMVFRSMLRQDVDFDPKNTTGALTTLRLANDAQAQVKAIGASRLAVIT 839
DB 832 AVCSELRIRARSTAFRTILRQDIAFDKESNTGALTSELTETKHLSCVSGVTGLTIL 891
QY 840 QNIANLGTGIIISLIYQWLTLALLAIPIAIAIVVEMKMLSGOALKDKKELEGAGKIA 899
DB 892 MTSTTLGAAIIIALAIGWKLALVCISVVPVLLACGFFRYFLMAQFQSRSLAYEGSANFA 951
QY 900 TEAIENFTVVSILTREQKFEYMYAQSLQVPRNSLRKAHIFGVSEFSTQAMMVFVSAGCF 959
DB 952 CEATSSIRTVASLTFRDREVDWEIYHAQDLDAQRTSLISVLRSLLIYASSQALVFCVALGF 1011
QY 960 RFQAYLVANEPFMDQDVLVLSAIVFGAMAVGVSSFPADYAKAKVAAHVIMIEKSPL 1019
DB 1012 WYGGTLLGHHEYDIFRPFVCFSEILFGAQAGTVPFSPADMGAKNAABEFLRFDKRPQ 1071
QY 1020 IDTSYPHGLKNTLEGNTVFNVEVFNTPDPDIPVLOGLSLEVKKGOTLALVSGSGCGKS 1079
DB 1072 IDNMSEGEKETVEGETEFRNVHRYTPRPEQPVLRGLDLTVRPGQIVALVGPSPGCGKS 1131
QY 1080 TVVOLLREYDPLAGSVLIDCKEIKHLNVQWLRAHLGLVISOEPTLIFDCSAENIAYGDSN 1139
DB 1132 TTIALREYDAIAGSILVDGKIDSKLNISYRSLVISOEPTLYQGTIKENILLGIVE 1191
QY 1140 RVVSHETIMQAKEANIHFIETLPEKYNTRVGDGKQOLSGGQKQRTAIARALVRQPHIL 1199
DB 1192 DDVPEELIKACKDANIYDFIMSLPEGENTVVGSGMGLSGGQKQRTAIARALLDRPKIL 1251
QY 1200 LLDDEATSDALTESEAVVQVALDKARKGRRTIVIAHRLSTVRNADVIAGFDDGVIVKEGNH 1259
DB 1252 LLDDEATSDALTESEAVVQVALDKARKGRRTIVIAHRLSTVRNADVIAGFDDGVIVKEGNH 1311
QY 1260 QQLLACKGYTFSMVSVOQAGAK 1280
DB 1312 SELVOKKRYEYELVNLQSLGK 1332

RESULT 9
US-09-328-320-2
; Sequence 2, Application US/09328320
; Patent No. 6228615
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene at/d of
; TITLE OF INVENTION: Multiple Drug Resistance
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/328,320
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/996,545
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Webster, Thomas D.
REGISTRATION NUMBER: 39,872
REFERENCE/DOCKET NUMBER: X-11766
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3334
TELEFAX: 317-276-2763
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1334 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-328-320-2

Query Match 38.3%; Score 2481; DB 4; Length 1334;
Best Local Similarity 42.0%; Pred. No. 2.7e-225;
Matches 538; Conservative 223; Mismatches 458; Indels 62; Gaps 8;
QY 24 KNEKKEKPTVTFAMFYSNWLRLMLVGTMAIHHGAALPLMLVFGNMTDSFANAG 83
DB 90 KQLEIKWNISFFGLWRATKMDLIMVISTICAIAAASFQRIMLV----- 137
QY 84 ISRNKTFPVIINESITNNTHQFHNLEEMTYAYIYSGIGAGVLVAAYIQVFWCLAAG 143
DB 138 -----QISYDEFYDELKFNVLVYVIGIGFVTVYVSTVGTGYTGE 178
QY 144 ROILKIRKQFPAHMRQIGFVHDVHGLNRLTRTDVSKINEGIDKIGMFFHSIATFF 203
DB 179 HATQIRIYELISLRQNGIFDLGAGEVTRITADTNLTQDISEKVGSLTALATFV 238
QY 204 TGFIVGFRGKLTIVILAISPVGLSAAIWAAILSSFTDKELLAYAKAGAAVEELAI 263
DB 239 TAFITAYVKYKWLALICSTTVALVLTWGGGQFIIKYSKSLDSYGAGGTVAEEVSSI 298
QY 264 RTVAFGQKKELERYNNKLEAKGIGIKKAITANISIGAAFLLIYASALAFWYGTSLV 323
DB 299 RNATAFGTQDLAKQYEVHLDAAEKWCTKQIVMGFMIGAMGLMYSNGLGFWMGSRFL 358
QY 324 LSSEVSIGOVLTVPFVSLGAFSGQASPIEAFANARGAAVEIFKIIDNKPIDSYSKS 383
DB 359 VDGAVDGDILTULMALIGSFLGNVSPNAQFTNAVAAAIFCTIDRQSPLDPSNE 418
QY 384 GHKPDNIGKLEFNHFSYPSRKEVILKGLNKLKQSGOTVALYCNCGSGKSTTVQLMQ 443
DB 419 GKTLDHEGHEILRNKYHIYSPETVMDVSLSPACKTTLVLPSPGSGKSTVVGIVE 478
QY 444 RLDPDGMVICDQDITRNHRLREITGVWSQEPVLFATTIAENIRYG-----RENV 498
DB 479 RFYMPVRGTVLLDGHDKDLNRLWRQQLSVLSQEPVLFGTIYKNIHRLGLGTKEYNES 538
QY 499 MDE-----IEKAVKEANAYDFIMKLPNKFDLTLYGCAQLSGGQKQKORIAARALVNP 554
DB 539 EDKVRLEIENAKMANAHDFITLALPGYETNVGQGFLLSGGQKQKORIAARAVSDPKIL 598
QY 555 LDEATSDLTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADYIAGFDGGLVIVEKGNH 614
DB 599 LDEATSDLTSEGVVQAALERAABGRTTIVIAHRLSTIKTAHNVIVLVNGKTAEGQTH 658
QY 615 DELMKEGIVFKLVWTRGNEIE---LENATGESKSDALEMSPKDSGSS-----L 664
DB 659 DELVDRGGYRKLVEAQRINEQKEADALEDAEDLTNADIATKIKTASSASSDLGKPTT 718

QY 665 IKRSTRESIHAPOQDRKLTGKTEDLNENVPVPSFWRILK-----LNSTEMPYFVVGIFCA 720
DB 719 IDRTGTHKSVSSAILSKRP-----PETTPKYSMTLLKLFVASFNRPEIPTYMLIGLVFS 771
QY 721 IINGLOPAFSIIFSRIGIIFTRDEDEPETK-RONSNNFSLVLVLGLIISFIITFFLQGTFF 779
DB 772 VLAGGQPTQAVLYAKAISTLSLPEISOYSLKLRHDAFWLMEFVVGIIOTITQSTNGAAF 831
QY 780 GKAGEILTCLRIVMYFRSMLRDQVSWFDDPKNTGTALTRLANDAAQVKAIGSLAVIT 839
DB 832 AVCSELRIRARSTAFRTILRQDIAFDEKENSTGALTSTLETETKHLSSGVSVLTGIL 891
QY 840 QNIANLGTIIISLIYQWLTLLLAIVPIAIAAGVEMKMLSGQALDKKKELEGAGKIA 899
DB 892 MTSTTLGAAIIIALAIGWKLALVCISVVPVLLACGFYFYMLAQFQSKLAYEGSANFA 951
QY 900 TEAIENFRVTVSLTREQKFEYMAQSLQVYRNSLRKAHIFGVFSFSITQAMMYFSYAGCF 959
DB 952 CEATSSIRTVASLTRERDVWEIYHAQLDAQGRISLVLRSLLYASSQALVFFCVLGF 1011
QY 960 RFGAIVLANEEMNFQDVLVFSNAIVFGAMAVQVSPFADPYAKAKVSAAHVIMIEKSPL 1019
DB 1012 WYGGTLGHHEYDIFRFVFCSEILFQAQSGAGTFVFAPDMGKAKNAAAEFRLFDRKPQ 1071
QY 1020 IDSYPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLOGLSLEVKKGOTLALVSGSGGKS 1079
DB 1072 IDNWSSEGEKLETEGEIEFRNVHRYTRPEQPVLRGLDITVTPGQVVALVSPSGGKS 1131
QY 1080 TVQOLLERYDPLAGSLVDIGKEIKHLNVQMLRAHLGVSOEPIFDCSIAENIAYGDN 1139
DB 1132 TTIALLEFYDAIGSILVDGDKDISKINSYRSFSLVSOEPTLYQGTIKENILLGIVE 1191
QY 1140 RYVSHEEIMQAAKEANIHHFTETLPEKYNTVRGDKGTQLSGQKORAIARALVRQPHIL 1199
DB 1192 DDVPEELIKACKDANIYDFIMSPGEFTVVGSGMLSGGQKQKORIAARALLRDPKIL 1251
QY 1200 LDEATSDLTSEKVKVQEAALDKAREGRTCTIVIAHRLSTIQNADLIYVVFONGVKKEGTH 1259
DB 1252 LDEATSDLTSEKVKVQEAALDKAREGRTCTIVIAHRLSTIQNADLIYVVFONGVKKEGTH 1311
QY 1260 QOLLAQKGIYFSWVSQAGAK 1280
DB 1312 SELVQKGRYYELVNLQSLGK 1332

RESULT 10
US-08-612-521-2
; Sequence 2, Application US/08612521
; Patent No. 5786463
; GENERAL INFORMATION:
; APPLICANT: Peery, Robert B
; APPLICANT: Skatrud, Paul L
; APPLICANT: Thornnewell, Susan J
; TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patent Division/AEH
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,521
; FILING DATE:
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Hamilton, Amy E
REGISTRATION NUMBER: 33,894
REFERENCE/DOCKET NUMBER: X-9693
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3169
TELEFAX: 317-276-1294
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1408 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-612-521-2

Query Match 37.7%; Score 2441; DB 1; Length 1408;
Best Local Similarity 41.2%; Pred. No. 1.8e-221;
Matches 540; Conservative 256; Mismatches 450; Indels 64; Gaps 19;

Qy 14 NEWKMGKK----SKNKKKKK-----PTVSTFAMFRYSNWLDRMLVMTGVTWAAIIHGAAL 65
Db 111 DFLSKRKKKEERKKKEKEASVLPVPSFFALPFAAPLEIIAMVLGLVLAAGAQCQ 170
Qy 66 PLMLVFGNMTDSFANAGISRNKTFPIINESITNTQHFHINLEEMTY----AYYS 121
Db 171 PLMTLIFGRLTTSFTNYAVIANQ-----ISOGGLTPETSAALQAADDLTKQSHNALYLM 226
Qy 122 GIGAGVLAAVIOVSFCLAAAGRQILKIRKQFFHAIMRQIEGWFVDVHDVGEINTRLTDDV 181
Db 227 AIGIGMFLAWLWYFHWNVGELNSKRIRERYLAALVRLQEIAYFDDLGAGEVATRIQTD 286
Qy 182 SKINEGIDGKIMGFHPSIATFTFGVTRGWKLILVILASPVLGLSAAIWAKLSF 241
Db 287 HLVBQSTSEKVALVQYAGTFVCGFLAFVRSPLAGALVSLIPVIMLCGGIMMTAMAKY 346
Qy 242 TDKELLAYAKAGAAVEELAAIRTVIAFGGKKELERYKNLEEAAGIGIKKAITANISI 301
Db 347 GTAALDHIAKAGSLAEVIGSIRTVAFGKREKILGDKFADHIOQSKIVGRKGSIFEGFGL 406
Qy 302 GAAFLIYASALAFWGTSLVLSSEYSIGQVLTVPFVSVLIGAFSGQASPSFEAFANAR 361
Db 407 SIMFFVIYAALAFFYGGILVNSQADSGIVINVFMSILIGFSMAMLAPELAAVTKAR 466
Qy 362 GAAYEIFKLIIDNKSIDYSKSGHKPNIKGNLEKFNHFSYPSRKEVKILKGLNLKVQS 421
Db 467 GAAAKLFATIDRVPADISASBEGPKDGLRGEISFENVKHYPSRPSIPLKGFITFEA 526
Qy 422 GOTVALVNSGCGKSTTVQLMORLYDPTDGMVCIDGQDITRVNHLREITGVVVSQEPVL 481
Db 527 GYTFALVNSGSGKSTTVVSLIERFYDVPVGVVVKLDGRDIRSLNLNLRQOIGLVSQEPTL 586
Qy 482 FATTIAENTRYG-----RENTWME----IEKAVKEANAYDFIMKLPNKFTTLVGERGAQ 532
Db 587 FGTVTRGNVEHGLIGSRYESNASLEEKFEVLVKKACVDANAHNFIMKLPQGYDTWVGERGML 646
Qy 533 LSGGQQRITAARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLS 592
Db 647 LSGGQQRVAIRAIVSDPRILLDEATSDALDTESEGIVQDALDKASGRRTTIIAHLRS 706
Qy 593 TVRNADVIAGFDGVTVEKGNHDELM-KEKGIYFKLVMTQTRGNEIELENATGESKSESD 651
Db 707 TIRDAIRIVYMGGEVLEQSGSHNDLANENGFPYQLVNNQKLAQE-----AAALQVDD 761
Qy 652 ALE-----MSPKDSGLSLIKRRSPRSIHAPOQDRKLGTKEDL-NENYPPVPSFW 700
Db 762 DIEDPDADVIGGSSPMQEKDQLHRAVYTRGSLASIAMDDIOAKRAEEVAGEDKIPSSFG 821
Qy 701 ---RIKLNSTEVPYFVVGIFCAITNGGLQAFSIFRSIGITFRDDEPTEKRONSMF 757
Db 822 LYARLLRMNSADRFIVIAAICAGMYPVPSLILFKALSDF-EIQDPAELRHALSRS 880
Qy 758 SVFLVLGIIISFTFFLOGFTGKAGEILTKLRYMVFPSMLRQDVSWFDPDPKNTTGALT 817

Db 881 ALWYFITALAAAFVIFFOAGFSRAGWDLNGLVRKKLFTATLRHDIWFDEERNSTGAVT 940
Qy 818 TRLANDAAQVGAIGSLAVITQNIANLNGIITISLYGWQLTILLIAIVPIITAIAGVVE 877
Db 941 SNLADPOKQVQGLFPTLGTVVQSCATLIGGCIIGLCYGPALLIGIACIPILVSGGIYR 1000
Qy 878 MKMLSGQALKDK--KELEGA--GKIATEAIENFRVTVSLTREQKFEYMYAOSLOVP----Y 930
Db 1001 LKVV---VLKDRMKLHAASAHLASEAAGAVKTVASLTREKDVRRYISALKAPMKLNF 1057
Qy 931 RNSLRKAHIFGVSFITQAMMYFSYAGCFRGAVLVANEFMNFQDVLVFSVAVFGAMAV 990
Db 1058 RTSIKSQCLFAAS----QGLTFCIIALFYIIGALWIDAKYSTASTFVYVLSIVFASIQ 1113
Qy 991 GOVSSFPADYAKAKYSAAHVIMIIEKSPIDISYSPHG--LKPNTLEGNTFNVFVNYPT 1048
Db 1114 GNVTTFVPDASKANSASASIFPSIDNEPAINAESNEGVLDHKKHVGHVRIEGVHFRIPT 1173
Qy 1049 RPDIPVLOGLSLEVKKQGTALVSGSGGKSTVVOLEFFYDPLAGSVLIDGKEIKHLNV 1108
Db 1174 RGVVRLRLTIDVPAGTVYVALVSPGCGKSTTIQMLERFYDPLAGRVTLGDIKELNL 1233
Qy 1109 QWLRAHLGVSOEPIILFDCSIAENIAYGDSRV--VSHEEIMQAAKEANIHHFIETLPEK 1166
Db 1234 ASYRSOISLVSQEPTIYAGTIFRINILLGANKPIEEVTQDEIDAACKDANIYDFIVSLPDG 1293
Qy 1167 YNTRVSDGKTQLSGGQKQRIATARALVRQPHILLDEATSDALDTESEKVVQVQALDKAREG 1236
Db 1294 FDIETVGGKSQLSGGQKQRIATARALIRNPKVLLDEATSDALDSEKVVQVQALDKAAG 1353
Qy 1227 RTCIVTAHLRSTONADLIVVFQNGKVBKRGHGTQOALLAQKGIYFSMVSVQ 1276
Db 1354 RTTIIAHLRSSIQHSDRIYFSEGRVAEHTGHOELLAKKGGYELVQM 1403

RESULT 11
US-08-395-246C-2
Sequence 2, Application US/08395246C
Patent No. 5773214
GENERAL INFORMATION:
APPLICANT: Peery, Robert B.
APPLICANT: Skatrud, Paul L.
TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
TITLE OF INVENTION: ASPERGILLUS FLAVUS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/395,246C
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: plant, Thomas G.
REGISTRATION NUMBER: 35784
REFERENCE/DOCKET NUMBER: x9683
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
TELEFAX: 317-277-1917
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1307 amino acids
TYPE: amino acid
TOPOLOGY: linear

Db	978	CWALGFVYGGSLGHGEYSLFQ-FYVCFSEVIFGAAGCTVPFSDMPDMGKAKHAAREFKR	1036
QY	1013	IIKESPLIDSYSPLGKPNLTLEGNVTFNENVTYTRDPIPVQLGSLSEVKKGQTLALVG	1072
Db	1037	LFSSDTHMASRS-KGVPVTSNRLGVDFEFDVSPRYSRLEQPTLRHLNITIKPGOFVALVG	1095
QY	1073	SSGGCKTTVVOLLERFYDPIAGSVLDGKEIFKHLNVQWLRHAHLGVISOEPTLPDCSTAEEN	1132
Db	1096	ASGGCKTTHALLERFYDPLKGGVYVDGKNIIITLEWSSVRSYSHLALISOEPTLPQGTIREN	1155
QY	1133	IAYGDSRVVSHHEIMQAKEANIIHFETLPEKNTYRVGDKGTQLSGGQKQRIATARAL	1192
Db	1156	ILGSNTPHVTDFVLVCKACKDANIVDFILSLPQSGENTIVGNKGWMLSGGQKQRIATARAL	1215
QY	1193	VRQPHILLDEATSDLTSESKVQOEALDKAREGRTICIVIAHRLSTTONADLIIVFPQNGK	1252
Db	1216	IRNPKILLDEATSDLTSESKRVQOALDAARGTTITAVAHRLSTIQRIADLIIVLDQGE	1275
QY	1253	VKEHGTHQOLLAQKGIYFSWSVQ	1276
Db	1276	VWESGTHRELLKRGYYELVHLQ	1299

Query Match	35.3%	Score 2287	DB 1	Length 1307
Best Local Similarity	39.6%	Pred. No. 6.1e-207		
Matches 509	Conservative 251	Mismatches 454	Indels 70	Gaps
QY	34	VSTFAMPRYSNWLDRLYMLVGTWAAIIHGAALPLMWLVFGNNWTSFANAGISRNRKTFPVI	93	
Db	45	IGPGIYRYATRWDAVLLFGSALAAAGGAALPLFTVLFGRLT		
QY	94	INESITNTQHFHNLHEEEMTYAYYSIGSAGVLVAAYIQVSEWCLAAAGROILKIRKQF	153	
Db	94	ATHRIT-----YDFHHELTKNVYFIYLGAAEFVAILATVGIYTGHRVVOIKREY	147	
QY	154	FHAMTROIQWFDVHDVGLNRLTRDDVSKINEGIDKIGMFHHSIAFTFTGIVGFTGR	213	
Db	148	FQALLQRONTAFPTDLAGETITRTADNLQDGSIEKVLGALFGLSTFTVTAFTIAIKN	207	
QY	214	WKLTVLILASPVLIGLSAAIMAKILISSTFKDELLAYAKAGAAVEVLAAARTVIAFGGQK	273	
Db	208	WKLALICSALLALLTIMGCCSTMLIFSKKALEYQQRGASMAEDILDSIRTVAAANAQGE	267	
QY	274	KELERYKNLEAKGIGIKKATANISITGAAPFLIYASYALAFWIGTSLVLS-----SEYSTG	331	
Db	268	TLARKYTESHLKDAEPGMKSRKIFAIMVGALLCTWLYNLGIFGFWGSRFLVEGINSIKAG	327	
QY	332	QVLTVFVSLIGAFSIGTQASPIEAFANARCAAYEIFKIIDNKPISDTSYKSGHRPDNIK	391	
Db	328	DVLTIMMALLIGSYNLGNVAPNGQALSDVAASAKLYGTIDROSPDILDALSDQGLTFEVR	387	
QY	392	GNLEPKNVHFSYPSRKEVKILGLNLKVSQGTVALVNGSCGKSTVOLMORLYDPTDG	451	
Db	388	GNIVLQNTIRHVPYSPREYTVADHLSCTVPAGKTTAFVGPSPGCKSTIISLLEFRFDPVAG	447	
QY	452	MYCIDGQDRIITNVRLREITGVVVSQBPVLFAITIAENIRYG-----RENV--TMDEI	502	
Db	448	TIMLDGHDIQTLNLRLRQMSLVVSQBPRLFAITIAENIRYIGISRRFEKESTYEIRRV	507	
QY	503	EKAVEANAYOPTIMKLPNKFDTLVGERGAOLSGQKQRIATAIALVRNPKILLDEATSA	562	
Db	508	EAARMANAHDFINALPNGVDYNTI--ESFSLSGQKQRIATAIAIKPKILLDEATSA	565	
QY	563	LDTESEAVVQVALDKARKGRTTIVIAHRLSTVFNADVIAGFDGCVIVKEGNHDELMKEG	622	
Db	566	LDTKSEKLVQALDKASKGRTTIVIAHRLSTIQKAYNIIVLANGQIVPEQGPHEHLMDRG	625	
QY	623	YFKLVMTQTRGNEILEENATGESKESDAL-EMSPK-----DSGS-----SL	664	
Db	626	LYCDMV----EAHEIK-KRYSRKYRSQLLTNLSPKHPNMTFFFDKDPGDDESIDYSI	680	
QY	665	IKRSTRSRTHAPOGDKLGTREDLNNVPPV-----SFRILK-----LNSTEMPYFVV	715	
Db	681	LSDASDGLHT--GEQRPVSRMSLSHLMQVKEEAYSFWTLFELFASFNRPWPFLLL	738	
QY	716	GIFCAILINGLOPAFSIFRIIGITPRD--EDPTEKQNSNMFSLVPLVLGILSIFTF	773	
Db	739	GLCASILAGGIQPSQAVLFKAVKSTLSLPLEYPKL-RHDANFCLMFLMIGIVSLVLYS	797	
QY	774	LOGFTFGKAGEILTCLRMYVFRMSLRQDVSWFDDPKNTGALTTRLANDAAQVKGAIGS	833	
Db	798	VQGLTFAYSEKVMYRARSQAFRVILHIDISFDDQENITGALTATLSAGTKELTGISGV	857	
QY	834	RLAVITONIANLGTGIIISLIYGWOLTLILLAIIVPIIAAGVVKMKLSQALKDKKELE	893	
Db	858	TLGTILIVSVNLVAGSLVIGWKLALVCISAVPALLMCGFVRVMMLERFORRAKKAYQ	917	
QY	894	GAKKIATEALENFRVTVSTREQKEFYMYAQSLQVPVYNSLRKAHIFGVFSFSTIQAMFV	953	
Db	918	ESASSACEAASAIRTVVSLTMETEALQSYQAOLRQLKSDILPIVKSLLYASSQALPFF	977	
QY	954	SYAGCFRFGAYLVAN-EEFMFDVLLVFSIAIVFCAMAGVQSFADYAKAKVSAAHVIM	1012	

```

Db      978 CWAJFWGSGLLGHGEYSLFQ-FYYCFSEVIFGAQAAGTVPESHAPDMGAKAHAAREFKR   1036
QY      1013 IIEKSPILDSYSPHGLKPNTLEGNWTFNEVENYPTRPDIPVLQGLSLEVKKGTALVG     1072
       :    :    :    :    :    :    :    :    :    :    :    :    :    :    :
Db      1037 LFSSDTMHASRS-KGVPTVTSMRGLVEFRDVSYFSRLEQPILRLHLNLTIKPGQFVALVG   1095
QY      1073 SSGCGKSTVVQLLERYDPDLGSLVDGKEIKHLNVQWLRAHLGLGVSOEPIFLFCSDIAEN   1132
Db      1096 ASGGKSTTALLERFYDPLKGVVYVDGKNIIITLEMSSYVRSHLALISOEPTLFQGTIREN   1155
QY      1133 IAYDNRRVVSHEEIIMAAKEANIHHFIETLPEKYNTNRVGDGTQLSGGOKQRATARAL     1192
       :    :    :    :    :    :    :    :    :    :    :    :    :    :    :
Db      1156 ILLGSNTPHYTDDBLVKACKDANIDYFILSLPQGENTIVNGKGMWSGGOKQRATARAL     1215
QY      1193 VROPHILLDEATSALDTESEKVQEQALDKKAREGRTCTIVIAHRLSTIQNALDIWVFQNGK   1252
Db      1216 IRMPKILLDEATSALDSESKRVQQAALDAARGRTTTIAVAHRLSTIORADLIWVLDQGE     1275
QY      1253 VKEHGTHQOALLAQKGIYFMSWSVQ   1276
       :    :    :    :    :    :    :    :    :    :    :    :    :    :    :
Db      1276 VESGTHRELLRKRGYYELVHLQ   1299

RESULT 12
US-08-232-537-2
; Sequence 2, Application US/08232537
; Patent No. 5516655
; GENERAL INFORMATION:
; APPLICANT: Peery, Robert B.
; APPLICANT: Skatrud, Paul L.
; TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
; TITLE OF INVENTION: AUREOBASIDIUM PULLULANS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,537

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X912
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
TELEFAX: 317-276-1917
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
JS-08-232-537-2

```

Query Match          28.6%: Score 1850; DB 1; Length 1302;
Best Local Similarity 34.2%: Pred. No. 1.5e-165;
Matches 440; Conservative 257; Mismatches 484; Indels 106; Gaps 2
92 VINESITNTTQHFTHLNEEMTTYYVYGGIGAGVLVAAYIQVSWFLAAGRILKIRK 151
      ||| | | | | | | | | | | | | | | | | | | | | | | | | |
66 SSYLD-LFKLVNDAKSKAVLNAGILLATAACGPPIPTGYTFGQIIITSFPPP----- 115
43 SNWIDRLYLMLV-----GTMAAILHGAALPLMMLVFNGNMTDSFANAGISRNKTPP 91
      ||| | | | | | | | | | | | | | | | | | | | | | | | | |

```

Thu Nov 7 09:13:10 2002

Db 1153 YNIGLGAAPGQVLTDRDIEKICAKGIEHIFIMSLPEGYSTECGTNGSKLSGGQKQRIAVA 1212
QY 1190 RALVROPHILLDEATSDALDTESEKVVQEQALDKAREGTCIVIAHRLSTIQNADLIIVVFQ 1249
Db 1213 RALIRSPVLLDEATSDALDTESEKVVQEQALDKAREGTCIVIAHRLSTIQNADRIEVED 1272
QY 1250 NGKVVHEGTHOOLLAQKGIYFWSVSVQ 1276
Db 1273 DGRVVEVGHAEVLAQGGIYAGMVLQA 1299

RESULT 13
US-08-996-644-2
; Sequence 2, Application US/08996644
; Patent No. 5945324
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten
; TITLE OF INVENTION: Multiple Drug Resistance Gene atrC of
; TITLE OF INVENTION: Aspergillus Nidulans
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,644
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11765
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-2763
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1308 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-996-644-2

Query Match 28.4%; Score 1838; DB 2: Length 1308;
Best Local Similarity 35.2%; Pred. No. 2e-164;
Matches 436; Conservative 251; Mismatches 502; Indels 50; Gaps 22;

Db 116 -----EDVLRDLQIVGVACGYFIVTTGYAIAAGLTGKESIRRFRE 157
QY 152 QFFHAIMROEIGWFDVHVGELNTRLTDDVSKINEGIGDKTGMFPHSIAFTFGFIYVGF 211
Db 158 TLVERLLGLEQAVFDIKD-PDITNLLTEKIEAIOIGTSEKVGIFQISYFVAAFIVGFI 216
QY 212 RGWKLTP-LVILAISPVGLSAAIAWAKILSSFTDKELLAYAK-AGAVAEVLAARTVIAF 269
Db 217 LNAKLTGILFAAIVPLMALIVTGGSRITAKYT-KAATYETAAGRIAESAHAVKVVQAF 275
QY 270 GGOKKELRYNKLEBAKIGIKKAITANISIGAAFLIYASALAFWYGTSLVL-SSEY 328
Db 276 GMAENLSKEHYRLKLSARVAKSVSAFMLGLVFTYSAANALAFWEGSRLAASEGNS 335
QY 329 SIGOVITVFSVLIGAFSPISQASPSIEAFANARGAAYETFKITIDNKP-IDSYSKSHK- 386
Db 336 NAGTVAVVFLIIDAISFVVGQFPFLGSPATAAAGESVYEILNHPQSEINVIYSEAGQEA 395
QY 387 -PDNIKNLFEKVNHSYPSRKEVKILKGLNKLKVSQGTVALVNGSGCGKSTTVOLMORL 445
Db 396 TESDMKADLVFRNVTFFVPARTSARALEMSLLIKAGOMNAIVGTSGCGKSTLVSLRL 455
QY 446 YDPTDGMVICIDQDIRTINVRHIREITGVVQEPVLFATTIAENIRYG-----RENVTM 499
Db 456 YDISSGQLTIGSHDKIDFNVRSLKRYTALVDQDSVLFSGSVLENTSYGHELSLDDVVL 515
QY 500 DEIEKAVKEANAYDFIMKLNKPDTLVGERG-AOLSGGQKQRIARALVRNPKILLDE 558
Db 516 ERCTEAKAAN-LDFVDFLOGIHRIGNGGYTSLSGQGNQICLARALVKKPALLLDE 574
QY 559 ATSALDTESEAVQVALDK-ARKGRTTIVIAHRLSTVRNADVTAGFDGDVIVEKGNHDEL 617
Db 575 PTAALDANSEGLINDAVKSAATCTTVMVAHRLSTVSDSPNVLTMGAGKVIIEQGNHDEL 634
QY 618 MKEKIYFKLVTMOTRGNEIELENATGESKESDAL---EMSPKDGSSLLIKRRSRSI 674
Db 635 MOLEGAYFNLOAOQL-----NDADSSAEVSAANTSQVTPQKASKESDSAAASDDET 687
QY 675 HAPQODRKLTGKEDLNENVPVFSW----RILKLNTEWPFYVVGIFCAIINGGLOPAF 730
Db 688 VPPQAK-----KED--KPAKKAGFWKLLRLCLRLAKSDSPIIAGLAAASIVSGGII 739
QY 731 SIIFSRIGITFRDEDPETKRONMSVFLVLGIIISFTFFLQGFTEFGKAGEILTKRL 790
Db 740 AIVFGNLISLVNDELSPDF-RSRADLFSLFFILALIALFSYAGNGCCFGIVSSHFAKI 798
QY 791 RYVFRSMLRQDVSWFDDPKNTGTALTRLANDAAQVKAIGSRSLAVITONIAMLTGII 850
Db 799 QHISLASILRQDMQWFGS-QSVPSLMSLSSDAGQLACLSGVAIGTFTVCVSVITGGII 856
QY 851 ISLIYQWQTLTLLLAIVPIIAIAGVVEKMLSGQALKDKKELEGKATATEAIEENFTVV 910
Db 857 LAHVWAKIAVLLAAVPMITAGVRLVRLVLALESRSRAYSANDASIAABACGIFTIA 916
QY 911 SLTREQKFEVMAQSLQVPRNSLRKAHT---FCVSESIQAMMYFSYACCFRGLAYV 966
Db 917 SLGREGVSRASNAWKEPYDKGIRFTLTNTLLALSFSIT-----YFYVLAAYWMAKQV 972
QY 967 ANEFMFQDVLVLYSFAIVFAMAVGVSSFADPDYAKAKVSAHVIMIEKSPLI---DSY 1023
Db 973 RNTGYSQDFFIVLPALLFSAOSAGQIFSLSPMSRAGVARNVFGLDHQKPTIVVDVDAK 1032
QY 1024 SPGLKRPNTLE-----GNVTENEVFNYPTRPDIPVLQGLSLEYKKQQTAL 1070
Db 1033 QSGALPSTLSIPTLEDKASPSSGGWIEFKNVSLCYPSKQPHALQNVNISIRPGEFIAL 1092
QY 1071 VGSAGCKSTVWQVLLERFYDPLAGSVLIDGKEIKHLNWOHLRAHGLTVSQEPILFDCSIA 1130
Db 1093 VGPSACKSTLSLQRFYDPTAGSVLQDQDIREVAVPQHRGLRGVLPQEPDLPFGSLS 1152
QY 1131 ENIAYG-DNSRVSHSEIMQAANEANHHFETLPKYNTRVGDGKTQLSGGQKQRIARIA 1189

505 LFGSGVEDNIRYGLVGTWPENASREEQMERYQEAALKLAYAHEFISELTDGYDTLIGRGG 564
QY 532 QLSGGQKORIAARALVRNPKILLDEATSALDTESEAVQVALDKARKGRTTIVIAHRL 591
Db 565 LLSGGQKORVARSVQPKVLLDEATSALDPAETIVQKALDKAEGRTTIVIAHKL 624
QY 592 STVRNADVIAGDDGVYVEKGHDELMKEGKIYFKLVMTQTRG-NEIELENATGESKSES 650
Db 625 ATRKADNIVNMSKGHIVEGQTHESLIAKDGVIAGLVKIQNLAVNASAHNDVNEEGEGED 684
QY 651 DALEMSPKDSGSLIKRRSTRSIIHAPQODRKLGCTKEDLNENVPVVSF-----WRILKL 705
Db 685 VAL-----LEVETAVTRYPT-----SIRGRMNSIKDRDY-ENHKHMDMLAALAYLVREC 734
QY 706 NSTEWPFVYVGFCAIINGG--LQPAFSIIFSRIGIIFTRDEDPETKRQNSNMFVSLFLV 763
Db 735 PELKWAYLVVLGCL--GCGAMYPCQAILMSRVVEFTLSD--AMLDKDFVASMILV 789
QY 764 LGIISFITFFLOQFTFGKAGEILTKRLRYMVFMSLRQDVSFWFDDPKNTTGALTIRLAND 823
Db 790 LAAGCLCYLAVGYATNTIAQHLSHWFRRLILHDMRLQDIQDFREENTGALVSRIDSY 849
QY 824 AAQVKAIGSRLLAVITQNTANIGTIIISLYGWQLTLL--AIVPIIAGVVMKML 881
Db 850 PHAILELMGYNIALVIAVLQVTCGILALAFSWKGLVWVFGGIPPLVG-AGMYRIRVD 908
QY 882 SGOALKDKKELEGAGKATEAENFRYVLSLREQKFYMYAQSLQVYPRNSLRKAHIFG 941
Db 909 SRLDRQTSKKYGTSSSIASEAVNAIRTVSSLAETEETVLRRYTEELDHAVSSVKPKMAATM 968
QY 942 VFSITQAMMYFSYAGCFRGALVANEFMNODVLLVFSATVFGAMAGVQVSSFPADYA 1001
Db 969 ICFLGTQCIETYQALGFYWGKLVLSGETSMYSFFVALSVFFAGQASQALFQWSTST 1028
QY 1002 KAKVSAAHVIMITEKSPLI--DSYSPHLKPTLEGNTFNEVFNYPTRDPIDVLOGLSL 1060
Db 1029 KGINATNYIAWLHQLQTVRETENHDKPGS-GAPIAMDNVRSFYPLRPDAPILKGVNL 1087
QY 1061 EVKQGOTLALVSGSGCKSTVOLLERFYDPLAGSVLIDGKEIKHLNQLVWLAHLGIYSO 1120
Db 1088 KINKQGTAFVSGSGCKSTMIAMLERFYDPTGTSITIDASTLTDINPISYRNIVALVQO 1147
QY 1121 EPIIFDCSIAENIAGVNSRVSHSEETMOAAKEANIHFFIETLPEKYNTRVGDGKTQLSG 1180
Db 1148 EPTLFQGTIRNISLGDAVKSVDQIESALRAANAWDFVSSLPGQIYTPAGSGGSQLSG 1207
QY 1181 GOKORIAARALVROPHILLDEATSALDTESEKVVQVQALD-KAREG-RTCIVIAHRLST 1238
Db 1208 GORORIAARALIRDPKILLDEATSALDTESEKIVQKALEGAARDGDRLTVAVAHRLST 1267
QY 1239 IQNADLIYVFGKVKHEGTHQQLAOKGIVFMSVQOA 1277
Db 1268 IKDANVICVFFGGKTAEMGTHQELIYVGLGYRRMCEAOA 1306

RESULT 15
US-08-463-092B-9
; Sequence 9, Application US/08463092B
; Patent No. 5766880
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,092B
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Steeg, Carol Miernicki
REGISTRATION NUMBER: 39,539
REFERENCE/DOCKET NUMBER: Q1546
TELECOMMUNICATION INFORMATION:
TELEPHONE: (613) 545-2342
TELEFAX: (613) 545-6853
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-463-092B-9

Query Match 19.9%; Score 1290.5; DB 1; Length 327;
Best Local Similarity 38.5%; Pred. No. 1.2e-113;
Matches 310; Conservative 9; Mismatches 8; Indels 479; Gaps 1;
QY 412 LKGLNLKVSQGVVALVGNSSGCKSTTVQLMORLYDPTDGMVCIDGQDRTINVRHLREI 471
Db 1 LKGLNLKVSQGVVALVGNSSGCKSTTVQLMORLYDPTDGMVCIDGQDRTINVRHLREI 60
QY 472 TGVYSQBPVLFAITIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPKFDTLVGERGA 531
Db 61 IGVYSQBPVLFAITIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPKFDTLVGERGA 120
QY 532 QLSGGQKORIAARALVRNPKILLDEATSALDTESEAVQVALDKARKGRTTIVIAHRL 591
Db 121 QLSGGQKORIAARALVRNPKILLDEATSALDTESEAVQVAL-----164
QY 592 STVRNADVIAGDDGVYVEKGHDELMKEGKIYFKLVMTQTRGNEIELENATGESKSES 651
Db 165 -----164
QY 652 ALEMSPKDSGSLIKRRSTRSIIHAPQODRKLGCTKEDLNENVPVVSFWILKLNSTWP 711
Db 165 -----164
QY 712 FYVWGIFCAIINGGLQPAFSIIFSRIGIIFTRDEDPETKRQNSNMFVSLFLVLIISFIT 771
Db 165 -----164
QY 772 PFLQGTETGKAGEILTKRLRYMVFMSLRQDVSFWFDDPKNTTGALTIRLANDAAQVKAI 831
Db 165 -----164
QY 832 GSRLAVITQNTANIGTIIISLYGWQLTLLLAIVPIIAGVVMKMLSGQALKDKKE 891

Db 165 ----- 164
QY 892 LEGAGKIATEAIENFRTVWSITREQKEPYMYAQSLOQVPYRNSLRKAHIFGVFSITQAMM 951
Db 165 ----- 164
QY 952 YFSYAGCFRFGAYLVANEFMFQDVLVFSIAIVFGAMAVGVSSFAPDYAKAKVSAHVI 1011
Db 165 ----- 164
QY 1012 MIIEKSPIDSYSPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLOGLSLEVKKGOTLALV 1071
Db 165 ----- 164
QY 1072 GSSGCGKSTVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHLGIVSOEPILEFDCSTAE 1131
Db 182 GSSGCGKSTVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHLGIVSOEPILEFDCSTAE 1131
QY 1132 NIAYGDSRVVSHHEIMQAAKEANTHHFTETLPEKYNTRVGDGKGTQLSGGQKQRIATARA 1191
Db 242 NIAYGDSRVVSOEEIVRAAKEANTHHFTETLPEKYNTRVGDGKGTQLSGGQKQRIATARA 1191
QY 1192 LVRQPHILLDEATSAIDTESEKVVQ 1217
Db 302 LVRQPHILLDEATSAIDTESEKVVQ 327

Search completed: November 6, 2002, 18:30:38
Job time : 35 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: November 6, 2002, 14:24:36 : Search time 83 Seconds
(without alignments)
12663.443 Million cell updates/sec

Title: US-09-672-725C-1
Perfect score: 4279
Sequence: 1 g9agcgcgagtcggtggtgg.....caaaaaaaaaaaaaaaaaaaaaa 4279

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Issued Patents NA.*
- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
 - 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
 - 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
 - 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
 - 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
 - 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3341.2	78.1	4669	6	5206352-3
2	3333.2	77.9	4669	2	US-08-752-447-1
3	3324.4	77.7	4646	1	US-08-181-471-2
4	3315.6	77.5	4669	2	US-08-583-276-18
5	3270.6	76.4	4264	2	US-08-784-649A-1
6	3270.6	76.4	4264	2	US-08-784-649A-5
7	3239.6	75.7	6505	2	US-08-793-610-5
8	3239.6	75.7	9318	2	US-08-793-610-6
9	2705	63.2	4233	3	US-09-120-513-1
10	2705	63.2	4233	4	US-09-450-105-1
11	2118	49.5	2726	1	US-08-461-823-1
12	656	15.3	4047	2	US-08-612-734B-1
13	606.8	14.2	4002	2	US-08-996-545-1
14	606.8	14.2	4002	2	US-08-996-545-3
15	606.8	14.2	4002	4	US-09-328-320-1
16	586.2	13.7	4224	1	US-08-612-521-1
17	586.2	13.7	4800	2	US-08-612-734B-3
18	585.4	13.7	4800	2	US-08-395-246C-1
19	437.8	10.2	3924	1	US-08-996-644-3
20	303	7.1	3924	3	US-08-996-644-3
21	303	7.1	3924	3	US-09-352-552-3
22	303	7.1	3927	2	US-08-996-644-1
23	303	7.1	3927	3	US-09-352-552-1
24	226.8	5.3	3909	1	US-08-232-537-1
25	178.4	4.2	4403765	4	US-09-103-840A-2
26	168.6	3.9	2061	4	US-09-061-764A-17
27	159.4	3.7	2376	1	US-08-394-880B-1

28	156.8	3.7	6143	1	US-08-612-521-3	Sequence 3, Appli
29	152.2	3.6	1959	4	US-09-061-764A-4	Sequence 4, Appli
30	152.2	3.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
31	152.2	3.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
32	149.8	3.5	8370	2	US-08-488-706-1	Sequence 1, Appli
33	148.2	3.5	5120	3	US-08-772-270A-6	Sequence 6, Appli
34	144.4	3.4	2407	2	US-08-895-522-2	Sequence 2, Appli
35	144.4	3.4	2407	3	US-09-195-391-2	Sequence 2, Appli
36	142	3.3	1201	3	US-08-961-083-189	Sequence 189, App
37	141.8	3.3	7721	3	US-08-772-270A-14	Sequence 14, Appl
38	141.6	3.3	2244	4	US-09-061-764A-18	Sequence 18, Appl
39	140.6	3.3	3576	4	US-08-976-259-79	Sequence 79, Appl
40	139.2	3.3	28804	2	US-08-592-874-1	Sequence 1, Appli
41	139.2	3.3	28804	3	US-09-096-942-2	Sequence 2, Appli
42	139.2	3.3	28804	3	US-09-096-867-2	Patent No. 5206352
43	131.6	3.1	807	6	5206352-1	Sequence 4, Appli
44	120.8	2.8	535	4	US-08-858-207A-230	
45	118.6	2.8	5110	2	US-08-404-531B-4	

ALIGNMENTS

RESULT 1
5206352-3
; Patent No. 5206352
; APPLICANT: Roninson, Igor B.; Pastan Ira H.; Gottesman, Michael M.
; TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
; SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/622,836
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 892,575
; FILING DATE: 01-AUG-1986
; APPLICATION NUMBER: 845,610
; FILING DATE: 28-MAR-1986
; SEQ ID NO:3
; LENGTH: 4669
5206352-3

Query Match	78.1%	Score	3341.2;	DB	6;	Length	4669;
Best Local Similarity	87.6%	Pred. No. 0;					
Matches	3758;	Conservative	0;	Mismatches	488;	Indels	44;
Gaps	8;						
QY	1	GGAGCGGAGTGGGATGGATCCTGAAGGAGCGCCGTAAGGGAGCTGCAGAGAAACTT	60				
Db	409	GGAGCGGAGTGGGATGGATCCTGAAGGAGCGCCGTAAGGGAGCTGCAGAGAAACTT	60				
QY	61	CTGGAANAATGGGCAAAAAGTAAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	468				
Db	469	CTTTTAACTGAACATAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	528				
QY	121	CAGGTTTGAATGTTTCGCTATTCAAATGGCTGCTGATAGTTGTTATGTTGGTGGGAC	180				
Db	529	TGATTTTCAATGTTTCGCTATTCAAATGGCTGCTGATAGTTGTTATGTTGGTGGGAC	588				
QY	181	AATGGTGGCATCATCCATGGAGCTGCATCCCTCTCATGATGCTGTTTGGAAACAT	240				
Db	589	TTTGGTGGCATCATCCATGGAGCTGCATCCCTCTCATGATGCTGTTTGGAAACAT	648				
QY	241	GACAGATAGCTTGAAGATGAGGAAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG	300				
Db	649	GACAGATAGCTTGAAGATGAGGAAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG	702				
QY	301	TGAAGATATAGCAACATACACACATTTTCATCAACATCTGCAGGAGGAAATGACCAC	360				
Db	703	TAATAGAGTATATCAATGATACAGGCTTTCATGAATCTGGAGGAGGAGGAGGAGGAG	762				
QY	361	GTATGCCATTATACATGGGAGCGGTGCTGGCTGCTGCTGGCTGCTTACATCCAGGT	420				

Thu Nov 7 09:13:08 2002

Db 763 GTATGCCCTATTATTACAGTGGAAATTGGCTGGGGTGCCTGGTTGCGCTTACATTCAGGT 822
QY 421 TTCATCTCGTGGCTGGCAGCAGGACAGATACACTCAAAATTAGAAAACAATTTTTTCA 480
Db 823 TTCAATTTGGTGGCTGGCAGCTGGAAGACAAATACACAAATTAGAAAACAGTTTTTCA 882
QY 481 TCGTATCATCGACAGGAGATGGCTGGTTTGAGCTGCATGACGTGGGGAGCTTAACAC 540
Db 883 TCGTATAATGCGACAGGAGATAGGCTGGTTTGATGTGCACGATGTTGGGAGCTTAACAC 942
QY 541 CGGGCTCACAGACGATGTCCTCAAAATCAATGAAGGAATTTGGCACAATAATGGAATGTT 600
Db 943 CCGACTTACAGATGATGCTCTCAAGATTAAATCAAGTTATTAGGTGACAAAATTGGAATGTT 1002
QY 601 CTTTCACTCAATAGCAACAATTTTTCACCGGTTTTATAGTGGGGTTTACACGTGGTGGAA 660
Db 1003 CTTTCAAGTCAATGGGCAACAATTTTTCACCTGGGTTTATAGTAGAATTTACACGTGGTGGAA 1062
QY 661 GCTAACCCCTTGTGATTTTGGCCATCAGCCCTGTTCTTGGAGCTTTCAGCCGCCATCTGGC 720
Db 1063 GCTAACCCCTTGTGATTTTGGCCATCAGCTCTTCTTGGACTGTGACGTGCTGTGGC 1122
QY 721 AAGATACTATCTTCAATTTACTGATAAGAACTCTTGGCTATGCAAAAGCTGGAGCAGT 780
Db 1123 AAGATACTATCTTCAATTTACTGATAAGAACTCTTGGCTATGCAAAAGCTGGAGCAGT 1182
QY 781 AGCTGAAGAAAGTCTTAGCAGCAATCAGAACTGTGATGGCTTTGGAGACAAAAGA 840
Db 1183 AGCTGAAGAGTCTTGGCAGCAATTAGAACTGTGATGGATTTGGAGACAAAAGA 1242
QY 841 ACTTGAAGGTACACAAAATTTAGAAAGCTAAAGAAATTGGGATAAAGAAAGCTAT 900
Db 1243 ACTTGAAGGTACACAAAATTTAGAAAGCTAAAGAAATTGGGATAAAGAAAGCTAT 1302
QY 901 CACGGCAACATTTCTATTGGTCCGCTTTCTTATTGATCTATGCAATCATATGCTTGGC 960
Db 1303 TACAGCCAATATTTCTATAGGTGCTGCTGCTTCTGCTGATCTATGCTTGGC 1362
QY 961 TTTCTGATGAGGACCTTCTTGGTCCCTCCAGTGAATTTCTTATGGACAACTACTCAC 1020
Db 1363 CTTCTGATGAGGACCTTCTTGGTCCCTCCAGGGAATTTCTTATGGACAACTACTCAC 1422
QY 1021 TGCTTCTTTCTGATTAATTTGGGCTTTTAGTATGGACAGCATCCCAAGCATGCA 1080
Db 1423 TGTATCTTTCTGATTAATTTGGGCTTTTAGTATGGACAGCATCTCCAAGCATGCA 1482
QY 1081 AGCATTTCAACGCAAGAGAGCAGCTTATGAATCTTCAAGATAATTGACAAATAAAC 1140
Db 1483 AGCATTTCAAAATGCAAGAGAGCAGCTTATGAATCTTCAAGATAATTGATTAAGCC 1542
QY 1141 AAGCATTCACAGCTATTGGAAGAGTGGACATAAACACAGATAATATTAAAGGAATTTGA 1200
Db 1543 AAGTATTGACAGCTATTGCAAGAGTGGCACAACACAGATAATATTAAAGGAATTTGA 1602
QY 1201 ATTCAAAAATGTTCACTTCACTTACCTTTCGAAAAGATTTAAGATCTTAAGGGTCT 1260
Db 1603 ATTCAAAAATGTTCACTTCACTTACCTTTCGAAAAGATTTAAGATCTTGAAGGCC 1662
QY 1261 CAACCTGAAGTTTCAGAGTGGCAGACAGTGGCTGGTTGGGAACAGTGGCTGCGGGAA 1320
Db 1663 GAACCTGAAGTTTCAGAGTGGCAGACAGTGGCTGGTTGGGAACAGTGGCTGCGGGAA 1722
QY 1321 GAGCAGACCGTGCAGCTGATCAGAGGCTCTATGACCCCAACAGATGGCATGGCTGTAT 1380
Db 1723 GAGCACAACAGTCCAGCTGATCAGAGGCTCTATGACCCCAACAGATGGCATGGCTGTAT 1782
QY 1381 TGATGACAGGACATTTAGGACCATTAATGTAAGGCATCTTCGGGAAATTTACTGGTGGT 1440
Db 1783 TGATGACAGGATTTAGGACCATTAATGTAAGGCTTCTACGGGAATCATTTGGTGGT 1842
QY 1441 GAGTCAGGAGGCTGTGTTGTTGGCCACCAAGATAGCTGAAACATTTGCTATGGCCGGA 1500
Db 1843 GAGTCAGGAGGCTGTGTTGTTGGCCACCAAGATAGCTGAAACATTTGCTATGGCCGGA 1902

QY 1501 AAATGTCACCATGATGAGATTGAGAAAGCTGTTAAGGAAGCAATGCCTATGATTTAT 1560
Db 1903 AAATGTCACCATGATGAGATTGAGAAAGCTGTTAAGGAAGCAATGCCTATGATTTAT 1962
QY 1561 CATGAAACTACCTAATAAATTTGACACTCTGGTTGGAGAGAGGGGCCAGCTGAGTGG 1620
Db 1963 CATGAAACTGCTCATAAATTTGACACCTCTGGTTGGAGAGAGGGGCCAGCTGAGTGG 2022
QY 1621 TGGACAGAAAACAGAGATCGCATTTGCTGGGGCCTGGTTTCGCAACCCCAAGATCTTCT 1680
Db 2023 TGGCAGAGAGCAGAGAGATCGCCATTTGCACGTGCCCTGGTTTCGCAACCCCAAGATCTTCT 2082
QY 1681 GCTGATGAGGCAACGCTCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGGTGGCTCT 1740
Db 2083 GCTGATGAGGCAACGCTCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGGTGGCTCT 2142
QY 1741 GAATAGGCCAGAAAAGCGCGGACTACCATTTGATAGCTCATCGTTTGTCTACAGTTCG 1800
Db 2143 GAATAGGCCAGAAAAGCGCGGACTACCATTTGATAGCTCATCGTTTGTCTACAGTTCG 2202
QY 1801 TAATGCGGATGTCATTTGCTGCTTTGATGATGGAGTCAATTTGGAGAAAGGAATCATGA 1860
Db 2203 TAATGCTGACGTCATCGCTGCTTTGATGATGGAGTCAATTTGGAGAAAGGAATCATGA 2262
QY 1861 TGAACCTCATGAAGAGAGGCGCATTTACTTCAAACTTCTCACAATTCAGACAGAGAGAAA 1920
Db 2263 TGAACCTCATGAAGAGAGGCGCATTTACTTCAAACTTCTCACAATTCAGACAGAGAGAAA 2322
QY 1921 TGAATTTGAGTTAGAAAATGCCACTGGTGAATTTCCAAAAGTGAAGTGAAGTGGTGGAAAT 1980
Db 2323 TGAATTTGAGTTAGAAAATGCCACTGGTGAATTTCCAAAAGTGAAGTGGTGGAAAT 2382
QY 1981 GTCTCAAAAGATTTAGGCTCAGTTTAAATAAAGAGATCAACTCGCAGAGATATACA 2040
Db 2383 GTCTTCAAAATTTAGGATCCAGTCTAATAAAGAGATCAACTCGTAGGAGTGTCCG 2442
QY 2041 TGCACCAAGGCAAGCAGACAGAAAGCTTGGTACAAAAGAGACTTGAATGAGATGTACC 2100
Db 2443 TGGATCAAGACCCAGACAGAGAAAGCTTAGTACCAAGAGGCTCGGATGAAGATATACC 2502
QY 2101 TCCAGTTTCCCTCTGGAGGATTTCTGAAGCTGAACCTCAACTGAATGSCCTTATTTTGTGT 2160
Db 2503 TCCAGTTTCCCTCTGGAGGATTTATGAAGCTTAAATTTAATGATGSCCTTATTTTGTGT 2562
QY 2161 TGGTATATTTGTGCTATTATAAACGGAGGCTGCAACAGCAGATTTTCAATTAATTTTC 2220
Db 2563 TGGTATATTTGTGCCATTTAATGGAGGCTGCAACAGCAGATTTTCAATTAATTTTC 2622
QY 2221 AAGGATTTATAGGGATCTTTACCCGAGATGAGGATCTGAAACAAAACAGACAGAAATAGTAA 2280
Db 2623 AAGGATTTATAGGGATCTTTACAGATTTGATGATCTGAAACAAAACAGACAGAAATAGTAA 2682
QY 2281 CATGTTTCTGATTTGTTCTAGTCTTGGAAATTTATTTCTTTTATTTACATTTTTCCTCA 2340
Db 2683 CTGTTGTTTCACTATTTGTTTCTAGCCCTTGGAAATTTTCTTTTATTTACATTTTTCCTCA 2742
QY 2341 GGGCTTCACATTTGGCAAGCTGGGGAGATCCTCACTAAGCGGCTTCGATACATGGTTT 2400
Db 2743 GGGCTTCACATTTGGCAAGCTGGGGAGATCCTCACTAAGCGGCTTCGATACATGGTTT 2802
QY 2401 CAGATCCATGCTGACACAGGATGTGAGTGGTTTGTGATGACCTAAAAACACCACTGGAGC 2460
Db 2803 CCGATCCATGCTGACACAGGATGTGAGTGGTTTGTGATGACCTAAAAACACCACTGGAGC 2862
QY 2461 ATTGACAACCGCTTGGCAATGATGCGGCTCAAGTTAAAGGGCTATAGGTTCCAGGCT 2520
Db 2863 ATTGACTACCGCTCGCCAAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCT 2922
QY 2521 TCGTGTCAATCCAGAGATATAGCAAAATCTGGGACAGGCTATTTATATCTTAAATCTA 2580
Db 2923 TCGTGTAAATTAACCCAGATATAGCAAAATCTTGGGACAGGATTAATTTATCTTCACTA 2982

Thu Nov 7 09:13:08 2002

REFERENCE/DOCKET NUMBER: 95,1121
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-9808
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4669 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..424
FEATURE:
NAME/KEY: CDS
LOCATION: 425..4264
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 4265..4669
US-08-752-447-1

Query Match 77.9%; Score 3333.2; DB 2; Length 4669;
Best Local Similarity 87.5%; Pred. No. 0;
Matches 3753; Conservative 0; Mismatches 493; Indels 44; Gaps 8;
Qy 1 GGAGCGGAGGTCGGGATGGATCCTGAAGAGGCGCTAAGGGAGTGCAGAGAAGAACTT 60
Db 409 GGAGCGGAGGTCGGGATGGATCTTGAAGGGACCGCAATGGAGGAGCAAGAAGAA 468
Qy 61 CTGGAATAAGGCAAAAAGTAAAAAATGAGAAAGAAAGAAAGAAACCACTGTGAG 120
Db 469 CTTTTTAAACTGAACATAAAAAGTAAAAAAGATAAGAAAGAAAGAAACCACTGTGAG 528
Qy 121 CACGCTTGCATATGTTTCGCTATTCAAAATGGCTTGATAGTGTGTATATGTTGGTGGGAC 180
Db 529 TGTATTTCAATGTTTCGCTATTCAAAATGGCTTGACAAAGTTGATATGTTGGTGGGAC 588
Qy 181 AATGGCTGCCATCATCATCAGAGTGCACATCCCTCTCATGATGCTGTTTGGAAACAT 240
Db 589 TTTGGCTGCCATCATCATCAGAGTGCACATCCCTCTCATGATGCTGTTTGGAAACAT 648
Qy 241 GACAGATAGCTTGCAAATGAGAAATTTCAAGAAACAAACAACTTTCCAGTTATATAATTA 300
Db 649 GACAGATATCTTGCAAATGAGAAATTTAGAAATCTGATGCT-----CAAAACATCAC 702
Qy 301 TGAAGATATTACGAAACATACACAACTTTCAACCACTCTGGAGAGAAATGACCAC 360
Db 703 TAATAGAAGTGATATCAATGATACAGGGTCTTTCATGAATCTGGAGGAAGCATGACCAG 762
Qy 361 GTATCGCTATTATACATGGATCGGTGCTGGCGTGTGCTGGCTGCTGCTGTACATCCAGGT 420
Db 763 GTATCGCTATTATACATGGATCGGTGCTGGCGTGTGCTGGCTGCTGCTGTACATCCAGGT 822
Qy 421 TTCATTCTGGTCCCTGGCAGGAGGAGACAGATCTCAAAATAGAAACAAATTTTCA 480
Db 823 TTCATTCTGGTCCCTGGCAGGAGGAGACAAATACACAAATAGAAACAAATTTTCA 882
Qy 481 TCGTATCATCGCACAGGAGATTGGCTGTTGACGTGATGACGTGTGGGAGCTTTAACAC 540
Db 883 TCGTATATAGCACAGGAGATAGGCTGTTGATGTGCACCATGTGGGAGCTTTAACAC 942
Qy 541 CCGGCTTCACAGACGATGTCTCCAAATCAATGAAGAAATGGCGACAAATTTGAATGTT 600
Db 943 CCGACTTCACAGATGATGTCTCCAAATTAATGAAGAAATTTGTCACAAATTTGAATGTT 1002
Qy 601 CTTTCTACTCAATAGCAACATTTTCCCGGTTTATAGTGGGTTTACACGTGGTTGGAA 660
Db 1003 CTTTCTAGTCAATGGCAACATTTTTCACGTGGTTTATAGTAAATTTACACGTGGTTGGAA 1062
Qy 661 GCTTAACCCCTGTGATTTTGGCCATCAGCCCTGTTCTTGACATTTTACGCCGCTCTGGCC 720
Db 1063 GCTTAACCCCTGTGATTTTGGCCATCAGTCTGTTCTTGACTGTGCTGCTGCTGGCC 1122

Qy 721 AAAGATACTATCTTCAATTTACTGATAAAGAACTCTTGGCCCTATATGCAAAAGCTGGAGCAGT 780
Db 1123 AAAGATACTATCTTCAATTTACTGATAAAGAACTCTTAGCTGATGCAAAAGCTGGAGCAGT 1182
Qy 781 AGCTGAAGAAGTCTTAGCAGCAATCAGAACTGATTCCTTTGGAGGACAAAGAAAGA 840
Db 1183 AGCTGAAGAAGTCTTAGCAGCAATTAGAACTGTGATTTGGAGGACAAAGAAAGA 1242
Qy 841 AGTTGAAGGTTACAAACAAAATTTAGAAAGAACTTAAAGAAATTTGGGATAAAGAAAGCTAT 900
Db 1243 AGTTGAAGGTTACAAACAAAATTTAGAAAGAACTTAAAGAAATTTGGGATAAAGAAAGCTAT 1302
Qy 901 CACGGCCCAACATTTCTATTTGGTGGCGCTTCTTATTTGATCTATCATCATCTGCTGGC 960
Db 1303 TACAGCCAAATTTCTATAGTGTGCTTCTGCTGATCTATGATCTTATGCTGTGGC 1362
Qy 961 TTTCTGGTATGGACCTCTTGGCTCTCCAGTGAATTTCTTATTTGGACAAGTACTCAC 1020
Db 1363 CTTCTGGTATGGGACCACTTGGTCTCTCAGGGGAATTTCTATTTGACAAGTACTCAC 1422
Qy 1021 TGTCTTCTTTCTGTATTTAAATTTGGGCTTTTATTTAGTATTTGGACAGGATCCCAAGCATTTGA 1080
Db 1423 TGTATTTCTTTCTGTATTTAAATTTGGGCTTTTATTTAGTATTTGGACAGGATCTCCAAAGCTTTGA 1482
Qy 1081 AGCATTTGCAAAAGCGAAGAGGAGCAGCTTATGAATCTTCAAGATAATTTGACAATAAAC 1140
Db 1483 AGCATTTGCAAAATGCAAGAGGAGCAGCTTATGAATCTTCAAGATAATTTGATAATAAGCC 1542
Qy 1141 AAGCATTTGACAGCTATTTGCAAGAGTGGACATAAACCAGATAATATTAAGGGAATTTTGGGA 1200
Db 1543 AAGTATTTGACAGCTATTTGCAAGAGTGGGACAAACCAGATAATATTAAGGGAATTTTGGGA 1602
Qy 1201 ATTCAGAAATTTTCACTTACCTTCTCGAAAAGAAAGTTAAGATCTTAAAGGGTCT 1260
Db 1603 ATTCAGAAATTTTCACTTACCTTCTCGAAAAGAAAGTTAAGATCTTAAAGGGTCT 1662
Qy 1261 CAACCTCAAGGTTTCAGAGTGGGACAGAGTGGGCTGTTGGGAAACAGTGGCTGGGGAA 1320
Db 1663 GAACCTCAAGGTTTCAGAGTGGGACAGAGTGGGCTGTTGGGAAACAGTGGCTGGGGAA 1722
Qy 1321 GAGCACACCGTGCAGCTGATGCAGAGGCTATGACCCCAAGATGGCATGGCTGTGAT 1380
Db 1723 GAGCACACAGTCCAGCTGATGCAGAGGCTATGACCCCAAGATGGCATGGCTGTGAT 1782
Qy 1381 TGATGGACAGACATTTAGGACCAATAAATGTAAGGTTTCTACGGGAAATCATTTGTTGTT 1440
Db 1783 TGATGGACAGATTTAGGACCAATAAATGTAAGGTTTCTACGGGAAATCATTTGTTGTT 1842
Qy 1441 GAGTCAGAGGCTGTGTTGTTGCCACCAAGATAGCTGAAAACATTCGCTATGGCCGCGA 1500
Db 1843 GAGTCAGAGGCTGTGTTGTTGCCACCAAGATAGCTGAAAACATTCGCTATGGCCGCGA 1902
Qy 1501 AAATGTCACCATGATGATGATTTGAAAGCTGTTTAAAGAAAGCAATTCCTATGATTTTAT 1560
Db 1903 AAATGTCACCATGATGATGATTTGAAAGCTGTTTAAAGAAAGCAATTCCTATGATTTTAT 1962
Qy 1561 CATGAAACTACTATAAATTTGACACTCTGTTTGGAGAGAGAGGGCCCAAGCTGAGTGG 1620
Db 1963 CATGAAACTGCTCATAAATTTGACACCTGTTTGGAGAGAGAGGGCCCAAGCTGAGTGG 2022
Qy 1621 TGGACAGAAACAGAGAAATCGCCATTTGCTGGGCGCTTGGTTCGCAACCCCAAGATTTCTT 1680
Db 2023 TGGGCAAGAGCAGAGGATCGCCATTTGACAGTGGCGCTTGGTTCGCAACCCCAAGATTTCTT 2082
Qy 1681 GCTGATGAGGCAAGCAGCTCAGCTCTGGACACTTGAAGTGAAGCAGTGGTTGAGTGGCCCT 1740
Db 2083 GCTGATGAGGCAAGCAGCTCAGCTCTGGACACTTGAAGTGAAGCAGTGGTTGAGTGGCTCT 2142
Qy 1741 GGATAGGCCAGAAAGGCCGAGCTTACCATTTGTGATAGCTCATCTGTTGTCTACAGTTGCG 1800
Db 2143 GGATAGGCCAGAAAGGTCGGACCAACCATTTGTGATAGCTCATCTGTTTGTGCTACAGTTGCG 2202

Db 4356 ATGAAGAGGATCTGTTTAAACATTTCCCTCAGTCAAGTTCAGAGTCTTC----- 4403
QY 4021 AGCTTTCAGATTTTATAAAGGAACAA--AAGAAACATTTATCTGATGGAATAAAAT 4078
Db 4404 -----AGAGACTTCGTAATTAAGGAACAGAGTGAGAGACATCATCAAGTGGAGAGAAAT 4458
QY 4079 ACTGGTGTAAATGGAATATAAAATATAGAGTAATTCAAAGTAGATTTTGTTAATAAT 4138
Db 4459 CATAGTTTAACTGCAATATAAATTTATAACAGAAATTAAGTAGATTTTAAAGATAAA 4518
QY 4139 TGTATAATTTTGTATATATTT-----ATTGTAACTTACTGCTTTTGTGAAAGAT 4190
Db 4519 ATGTGTAATTTTGTATATATTTTCCCATTTGGACTGTAACTGCTTGCCTGCTAAAGAT 4578
QY 4191 TATAGAAGTGGTAAGGACTG--AATGTTTGAATAAAGTGTGACTATATAATAAACTAA 4249
Db 4579 TATAGAAGTAGCAAAAGATTGAAATGTTTGCATAAAGTG---TCTATAATAAACTAA 4635
QY 4250 ACTTTTATATCAAAAAAAAAAAAAA 4279
Db 4636 ACTTTCATGTGAAAAAAAAAAAAA 4665

RESULT 3
US-08-181-471-2
; Sequence 2, Application US/08181471
; Patent No. 5641508
; GENERAL INFORMATION:
; APPLICANT: Li, Lingna
; APPLICANT: Lishko, Valeryi K.
; TITLE OF INVENTION: METHOD FOR DELIVERING BENEFICIAL
; TITLE OF INVENTION: COMPOSITIONS TO HAIR FOLLICLES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas Fitting
; STREET: 12526 High Bluff Drive, Suite 300
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92130
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 13-JAN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/041,553
; FILING DATE: 02-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: ANT0029P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-792-3680
; TELEFAX: 619-792-8477
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4646 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 425..4267
; US-08-181-471-2

Query Match 77.7%; Score 3324.4; DB 1; Length 4646;
Best Local Similarity 87.6%; Pred. No. 0;
Matches 3740; Conservative 0; Mismatches 486; Indels 44; Gaps 8;
QY 1 GGAGCGGAGGTGGGATGGATCCTGAAGGAGCGCCGTAAGGGGAGTGCAGAGAGAAGAACTT 60
Db 409 GGAGCGGAGGTGGGATGGATCCTGAAGGGAGCCCAATGGAGGAGCAAGAAAGAAGAA 468
QY 61 CTGGAATAATGGCAAAAAAAGTAAAAAAATGAGAGAAAAAGAAAAACCAACTGTCTAG 120
Db 469 CTTTTTAAACTGAACAATAAAGTGAAGAAGATAAGAGAAAAAGAAACCAACTGTCTAG 528
QY 121 CACGCTTTCAATGTTTCGCTATTCAAAATGGCTGTAGTGTATATATGTGTGGGGAC 180
Db 529 TGTATTTTCAATGTTTCGCTATTCAAAATGGCTGTAGTGTATATATGTGTGGGGAC 588
QY 181 AATGGCTGCCATCATCATGAGCTGCACCTCCCTCATGATGCTGCTTTTGTGGAACAT 240
Db 589 TTTGGCTGCCATCATCATGAGCTGCACCTCCCTCATGATGCTGCTTTTGTGGAACAT 648
QY 241 GACAGATAGCTTTGCAAAATGCAAGAAATTTCAAGAAAAACAAACATTTTCCAGTTATAATA 300
Db 649 GACAGATATCTTTGCAAAATGCAAGAAATTTTGAAGAATCTGTATGTCA-----AACATCAC 702
QY 301 TGAAGATATTACGAACAATACACAACATTTTCATCAACCATCTCGAGGAGAAATGACCAC 360
Db 703 TAATAGAAGTGAATATCAATGATACAGGCTTCTTCATGAATCTGGAGGAAGACATGACCAG 762
QY 361 GTATGCCCTATTATACAGTGGGATCGGTGCGTGTGCTGCTGCTGCTGCTTACATCCAGGT 420
Db 763 GTATGCCCTATTATACAGTGGGATCGGTGCGTGTGCTGCTGCTGCTGCTTACATCCAGGT 822
QY 421 TTCTATCTGCTGCTGCGCAGGAGGACACATCTCAAAATTAGAAATAGAAACAAATTTTCA 480
Db 823 TTCAATTTTGGTGGCTGCGCAGGCTGGAAGACAAATACACAAAAATAGAAACAGTTTTC 882
QY 481 TGCTATCATCGACAGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 883 TGCTATAATGCGACAGGAGATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 942
QY 541 CCGGCTCACAGACGATGTCTCCAAATCAATGAAGAAATTTGGCCACAAAAATGGAATGT 600
Db 943 CCGACTTACAGATGATGTCTCTAAGATTAATGAAGTTTATGCTGACAAAAATTTGAATGT 1002
QY 601 CTTTCACCTAATAGCAACAATTTTACCGGTTTATAGTGGGTTTACACGTTGTTGAA 660
Db 1003 CTTTCAGTCAATGGCAACAATTTTACCTGGGTTTATAGTAGGATTTACACGTTGTTG 1062
QY 661 GCTAACCCCTGTGATTTTGGCCATCAGCCCTGTTCTTGGACTTTTCCAGCCGCTATCGG 720
Db 1063 GCTAACCCCTGTGATTTTGGCCATCAGTCCTGTTCTTGGACTGTGCTGCTGCTGCTG 1122
QY 721 AAAGATACCTATCTTCAATTTACTGATAAGAACTCTTGGCTATGCAAAAGCTGGAGCAGT 780
Db 1123 AAAGATACCTATCTTCAATTTACTGATAAGAACTCTTGGCTATGCAAAAGCTGGAGCAGT 1182
QY 781 AGCTGAAGAAGTCTTAGCAGCAATCAGAACTGTGATTTGCTTTGGAGGACAAAAAGAA 840
Db 1183 AGCTGAAGAAGTCTTAGCAGCAATCAGAACTGTGATTTGCTTTGGAGGACAAAAAGAA 1242
QY 841 ACTTGAAGGTACACAAAAATTTTAGAAGAGCTAAGGAATTTGGATTAAGAAAGCTAT 900
Db 1243 ACTTGAAGGTACACAAAAATTTTAGAAGAGCTAAGGAATTTGGATTAAGAAAGCTAT 1302
QY 901 CACGGCCCAACATTTCTATTGTTGGCTCTTCTTATTTGATCTATGATCATATGCTCTGCG 960
Db 1303 TACAGCCCAATATTTCTATAGGTGCTGCTTTCTGCTGCTATGATCATCTTATGCTGCG 1362
QY 961 TTTCTGATATGGGACCTCTTGGTCTCTCCAGTGAATATTTCTATGGACAAGTACTCAC 1020
Db 1363 CTTTCTGATGGGACCCACCTTGTCTCTCAGGGGAATATTTCTATTTGGACAAGTACTCAC 1422
QY 1021 TGTCTTCTTTTCTGATTAATTTGGGGCTTTTACTATTGGACAGGCATCCCCAAGCATCGA 1080

Db 1423 TGTATTTCTTCTGTATTAATTTGGGCTTTTAGTGTGGACAGCACTCTCAACCATTTGA 1482
QY 1081 ACCATTTTCCAAACCAAGAGAGAGAGCTTATGAAATCTTCAAGATAATTTGACAATAAAC 1140
Db 1483 AGCATTTTGCAAATCAAGAGAGAGAGCTTATGAAATCTTCAAGATAATTTGATAATAAGC 1542
QY 1141 AAGCATTTGACAGCTATTTCGAAGAGTGGACATAAACCAAGATAATTAAGGAAATTTGGA 1200
Db 1543 AAGTATTGACAGCTATTTCGAAGAGTGGCAACCAACAGATAATTAAGGAAATTTGGA 1602
QY 1201 ATTCAAATAATTTTCACTTCAGTTTACCCTTCTCGAAAGAAAGTTAAGATCTTTAAAGGCTCT 1260
Db 1603 ATTCAGAAATTTTCACTTCAGTTTACCCTTCTCGAAAGAAAGTTAAGATCTTTGAAGGSCCT 1662
QY 1261 CAACCTGAAGGTTTCAGAGTGGGACAGAGTGGGCTGGTTGGGACAGTGGCTGGCGGAA 1320
Db 1663 GAACCTGAAGGTTTCAGAGTGGGACAGAGTGGGCTGGTTGGGACAGTGGCTGGCGGAA 1722
QY 1321 GAGCAGCAGCTGATGATGAGAGGCTCTATGACCCCAACAGATGGCATGTGTCTGTAT 1380
Db 1723 GAGCACAACAGTCCAGCTGATGAGAGGCTCTATGACCCCAACAGAGGGGATGTGAGTGT 1782
QY 1381 TGATGGACAGACATTTAGGACCATAAATGTAAAGCATCTTTCGGGAAATTAATCTGGTGTGT 1440
Db 1783 TGATGGACAGATATTAGGACCATAAATGTAAAGGTTTCTACGGGAAATCAATTTGTGTGT 1842
QY 1441 GAGTCAGAGGCTGTGTGTGTTCACACAGATAGCTGAAACATTCGCTATGCCCGCA 1500
Db 1843 GAGTCAGAGGCTGTGTGTTCACACAGATAGCTGAAACATTCGCTATGCCCGTGA 1902
QY 1501 AAATGTCAACATGGATGAGATTGAGAAAGCTGTTAAGAAAGCCCAATGCCCTATGATTTAT 1560
Db 1903 AAATGTCAACATGGATGAGATTGAGAAAGCTGTTAAGAAAGCCCAATGCCCTATGATTTAT 1962
QY 1561 CATGAAACTACCTAATAAATTTGACACTCTGGTTGGAGAGAGGGGCCAGCTGAGTGG 1620
Db 1963 CATGAAACTACCTAATAAATTTGACACTCTGGTTGGAGAGAGGGGCCAGCTGAGTGG 2022
QY 1621 TGGACAGAAACAGAGATGCCATTTGCTGGGCCCTTGGTTTCGCAACCCCAAGATTTCTTCT 1680
Db 2023 TGGCAGAGCAGAGATGCCATTTGACGTCGCCCTTGGTTTCGCAACCCCAAGATTTCTTCT 2082
QY 1681 GCTGGATGAGGCAACGCTACGCTTGGACACTGAAAGTGAAGCAGTGGTTTCAGGTGGCCCT 1740
Db 2083 GCTGGATGAGGCAACGCTACGCTTGGACACTGAAAGTGAAGCAGTGGTTTCAGGTGGCTCT 2142
QY 1741 GGATAGGCCAGAAAGGCGGACTACCATTTGTGATAGCTCATCGTTTGTCTACAGTTGG 1800
Db 2143 GGATAGGCCAGAAAGGCGGACTACCATTTGTGATAGCTCATCGTTTGTCTACAGTTGG 2202
QY 1801 TAATGCGGATGTCATTGCTGTTTGTGATGAGTGCATTTGGAGAAAGGAAATCATGA 1860
Db 2203 TAATGCTGAGTCTATCGCTGGTTTCGATGATGGAGTCATTTGGAGAAAGGAAATCATGA 2262
QY 1861 TGAATCATGAAGAGAGAGGCAATTTACTTCAAACTTGTCACAATGCAGACAGAGGAA 1920
Db 2263 TGAATCATGAAGAGAGAGGCAATTTACTTCAAACTTGTCACAATGCAGACAGAGGAA 2322
QY 1921 TGAATTTGAGTTGAGAAATGCCACTGGTGAATCCAAAGTGAAGTGAATGATGCTTGAAT 1980
Db 2323 TGAATTTGAGTTGAGAAATGCCACTGGTGAATCCAAAGTGAATTTGATGCTTGAAT 2382
QY 1981 GTCTCCAAAGATTTCAAGGTCAGTTTAAATAAAGAAAGATCAACTTCGAGGAGTATACA 2040
Db 2383 GTCTTCAATGATTTCAAGATCCAGTCTAATAGAAAAGATCACTCTGTAGAGTGTCCG 2442
QY 2041 TGCACACAAGGCCAAGACAGAAAGCTTGGTACAAAAGAGCACTTGAATGAGAATGTACC 2100
Db 2443 TGGATCACAAAGGCCAAGACAGAAAGCTTAGTACCAAAGGCTCTGGATGAAGATATACC 2502
QY 2101 TCCAGTTTCTTCTGGAGATTTCTGAGCTGAATCAACTGAATGATGCTTATTTTGTGTG 2160
Db 2503 TCCAGTTTCTTCTGGAGATTTATGAGCTAAATTTAACTGAATGGCCCTATTTTGTGTG 2562

QY 2161 TGGTATATTTTGTGCTTATTATAAACGGAGGCTTCGAACCAAGCATTTTCAATATATTTTC 2220
Db 2563 TGGTATATTTTGTGCCATTATAAATGGAGGCTTCGAACCAAGCATTTGCAATATATTTTC 2622
QY 2221 AAGGATATAGGATCTTTTACCAGAGATGAGGATCTGAAACAAAACGACAGAAATGAA 2280
Db 2623 AAAGATATAGGGGTTTTCACAGAAATGATGATCTGAAACAAAACGACAGAAATGAA 2682
QY 2281 CATGTTTTCTGTATTTCTAGTCTCTGGAATATTCTTTTATTACATTTTCTCTCA 2340
Db 2683 CTGTTTTTCACTATTTCTTACGCCCTTGGAAATATTCTTTTATTACATTTTCTCTCA 2742
QY 2341 GGGCTTCACATTTGGCAAGCTGGGAGATCCTCACTAAAGCGGCTTCGATACATGTTTT 2400
Db 2743 GGGTTTTCACATTTGGCAAGCTGGAGATCCTCAACAAAGCGCTCCGATACATGTTTT 2802
QY 2401 CAGATCCATGCTGAGACAGGATGTCAGCTGGTTTGTATGACCCCTAAACACCACTGGAGC 2460
Db 2803 CCGATCCATGCTGACAGAGATGTGAGTTGGTTTGTATGACCCCTAAACACCACTGGAGC 2862
QY 2461 ATTGCAACCAAGGCTTTGCCAATGATGCGGCTCAAGTTTAAAGGGGCTATAGGTTCCAGGCT 2520
Db 2863 ATTGACTACAGGCTCGCCAATGATGCTCTCAAGTTTAAAGGGGCTATAGGTTCCAGGCT 2922
QY 2521 TGGTGTCAATTTACCAGAAATATAGCAAAATCTTGGGACAGGCATTTATATCTTAACTA 2580
Db 2923 TGGTGTCAATTTACCAGAAATATAGCAAAATCTTGGGACAGGCATTTATATCTTAACTA 2982
QY 2581 TGGTTGGCAATTTAAACACTTTTACTCTTAGCAATTTGACCCATTTGCAATAGCAGGAGT 2640
Db 2983 TGGTTGGCAATTTAAACACTTTTACTCTTAGCAATTTGACCCATTTGCAATAGCAGGAGT 3042
QY 2641 TGGTGAATTTGAAATTTTGTCTGGCAAGCACTGAAAGATAGAAAGAGCTAGAAAGGAGC 2700
Db 3043 TGGTGAATTTGAAATTTTGTCTGGCAAGCACTGAAAGATAGAAAGAGCTAGAAAGGAGC 3102
QY 2701 TGGGAGATTTGCTACAGAGGCTCGAAACCTTCCGAATTTGTTTCTTTGACTCGGGA 2760
Db 3103 TGGGAGATTTGCTACAGAGGCTCGAAACCTTCCGAATTTGTTTCTTTGACTCGGGA 3162
QY 2761 GCAGAAAGTTGAATACATGATGCACAGATTTGCAAGTGTACCATACAGAACTCTTTTCAG 2820
Db 3163 GCAGAAAGTTGAATACATGATGCCTCAGATTTGCAAGTGTACCATACAGAACTCTTTTCAG 3222
QY 2821 GAAAGCACATCTTTTCGGGGTCTCAATTTCTATCACCCAGGCAATGATGATTTTCTCA 2880
Db 3223 GAAAGCACATCTTTTGAATTTACATTTTCTTCAACCAGGCAATGATGATTTTCTCA 3282
QY 2881 TGCTGGCTGTTTCCGGTTTGGTGCTTACTTGGTGCCATTTGGTGCCATGAGTTTTCAGGA 2940
Db 3283 TGCTGGATGTTTCCGGTTTGGAGCCTTACTTGGTGCCATTAACACTCATGAGCTTTGAGGA 3342
QY 2941 TGTCTTTTGTGATTTCTCAGCTTATTTGTTTGGTGCCATTTGGTGCCAGGTTCAGTTTC 3000
Db 3343 TGTCTGTTGATTTTTCAGCTGTTGTTTGGTGCCATGCGGCTGGGCAAGTTCAGTTTC 3402
QY 3001 ATTTGCTCTGACTATGCCAAGCCAAAGTATCAGAGCCCAAGTATGATGATGATGATG 3060
Db 3403 ATTTGCTCTGACTATGCCAAGCCAAAGTATCAGAGCCCAAGTATGATGATGATGATG 3462
QY 3061 AAAAGGCCCTCTGATTTGACAGCTACAGCCCTCAGGGCTCAAGCCCAATAGTTTGGAGG 3120
Db 3463 AAAAGGCCCTTGTGATGAGCTACAGCAGCGAAGGCGCTTAATGCGCAACATTTGGAAGG 3522
QY 3121 AAATGTGACATTTAATGAGGTCGTTTCAACTATCCCACTCGACAGACATTTCCCGTGTCT 3180
Db 3523 AAATGTGACATTTGTTGAGGTTGTTTCAACTATCCCACTCGACAGACATTTCCCGTGTCT 3582
QY 3181 CCAGGGCTGAGCCTCGAGGTGAAGAGGGCCAGACGCTGGCCCTCGTAGTAGCAGTGG 3240
Db 3583 TCAGGGCTGAGCCTCGAGGTGAAGAGGGCCAGACGCTGGCTCTGGTGGGACAGCAGTGG 3642

```
QY 3241 CTGTGGGAAGACACAGTGTGTTACAGTCTCTAGACAGCGCTCTATGACCCCTTGGCTGGTTC 3300
Db 3643 CTGTGGGAAGACACAGTGTGTTACAGTCTCTAGACAGCGGTCTACGACCCCTTGGCAGGAA 3702
QY 3301 AGTGTCTAATGTATGCAAGAGATAAAGCACCTGAAATGTCAGTGGCTCCGAGCACACCT 3360
Db 3703 AGTGTCTGTATGCAAGAAATTAAGGAGTGAATGTTCAAGTGGCTCCGAGCACACCT 3762
QY 3361 GGGCATCGTGTCTCAGGAGCCCATCTGTGTTGACGTGACAGATGTCGAGAGACATTTGCCCTA 3420
Db 3763 GGGCATCGTGTCTCAGGAGCCCATCTGTGTTGACGTGACAGATTTGCTGAGAACATTTGCTA 3822
QY 3421 TGGAGACACAGCCGGTCTGATACATGAAGAGATTAAGCAGGAGGAGGAGGAGGAGGAGG 3480
Db 3823 TGGAGACACAGCCGGTCTGATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3882
QY 3481 CATACACCACTTCTATCGAGACACTCTCCTGAGAAATACACACAGAGTAGGAGACAAAGG 3540
Db 3883 CATACATGCTTCTATCGAGTCACTGCTTAATAATATAGCACTAAAGTAGGAGACAAAGG 3942
QY 3541 AACCCAGCTCTGTGTGGCCGAAGACAGCGCATTTGCCATAGCTCGGCTCTTTTAGACA 3600
Db 3943 AACTCAGCTCTGTGTGGCCGAAGACAGCGCATTTGCCATAGCTCGGCTCTTTTAGACA 4002
QY 3601 GCCTCATATTTTCTGTTGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGT 3660
Db 4003 GCCTCATATTTTCTGTTGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGT 4062
QY 3661 TGTCCAAAGAGCCCTGGCAAGCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3720
Db 4063 TGTCCAAAGAGCCCTGGCAAGCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4122
QY 3721 CTTGTCCAGCATCCAAATGCAGATTAATAGTGGTGTTCAGAAATGCAAGTCAAGGA 3780
Db 4123 CTTGTCCAGCATCCAAATGCAGATTAATAGTGGTGTTCAGAAATGCAAGTCAAGGA 4182
QY 3781 GCATGGCACACATCAACAGCTGTGCGCCAGAAAGGATCACTATTTTCCATGCTGAGTGT 3840
Db 4183 GCATGGCACACATCAACAGCTGTGCGCCAGAAAGGATCACTATTTTCCATGCTGAGTGT 4242
QY 3841 CCAGGCTGGAGCAAGCGGTAGTGAAGTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3900
Db 4243 CCAGGCTGGAGCAAGCGGTAGTGAAGTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4302
QY 3901 ATTGTGTTAAACATGGCATTTAATCAAAAGTTAAAGGTGAGCACTTACTGGAAAGACT 3960
Db 4303 ATT--TGTTTAGATATGACATTTTATTCAAAGTTAA--AGCAACACATTACAGAAAT 4355
QY 3961 ATGTAGAACTACCTGTTTAAACATTTCTGCTGCAACTGAAGATCATTCACCAAGTTCAG 4020
Db 4356 ATGAAGAGGTATCTGTTTAAACATTTCTCAGTCAAGTTCAGAGTCTCTC----- 4403
QY 4021 AGCTTCAGATTTTATTAATTAAGGAGACCAA--AAGAACATTTATCTGATGGAATAAAT 4078
Db 4404 -----AGAGAGATTCGTAAATTAAGAGACAGAGTGAGAGACATCATCAAGTGAGAGAAAT 4458
QY 4079 ACTGCTGTTAATGCAATATATAAATATAGAGTAATCAAGTAGATTTTGTGTTAATAAT 4138
Db 4459 CATAGTTTAACTGCATTTAATTTTATTAACAGAAATTAAGTAGATTTTAAAGATATAA 4518
QY 4139 TGTATAAATTTGTTTATATTTT-----ATTGTAACCTTACTGCTTGTGCTGAAGAT 4190
Db 4519 ATGTGTAATTTTGTGTTATATTTTCCCATTTTGGACTGTGAACCTGCTGCTGCTTAAAGAT 4578
QY 4191 TATAGAGTGGTAAAGAGTCTG--AATGTTTGAATAAGGCTAGCTATATAATAAACTAA 4249
Db 4579 TATAGAGTGGTAAAGAGTATTGAAATGTTTGTGATTAAGTG----TCTATATAAACTAA 4635
QY 4250 ACTTTTATAT 4259
Db 4636 ACTTTCATGT 4645
```

```
RESULT 4
US-08-583-276-18
; Sequence 18 Application US/08583276
; Patent No. 5837536
; GENERAL INFORMATION:
; APPLICANT: McDonagh, Kevin T.
; APPLICANT: Nienhuis, Arthur
; APPLICANT: Tolstoshev, Paul
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
; TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
; TITLE OF INVENTION: SELECTION OF CELLS TRANSDUCED WITH SUCH GENES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi & Stewart
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: DM4.V2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/583,276
; FILING DATE: 05-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/332,444
; FILING DATE: 31-OCT-1994
; APPLICATION NUMBER: 07/887,712
; FILING DATE: 22-MAY-1992
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4669 bases
; TYPE: nucleic acid
; STRANDEDNESS: singular
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: Genomic DNA
US-08-583-276-18

Query Match 77.5%; Score 3315.6; DB 2; Length 4669;
Best Local Similarity 87.2%; Pred. No. 0;
Matches 3742; Conservative 0; Mismatches 504; Indels 44; Gaps 8;

QY 1 GGAGCGCGAGGTCGGGATGATCCTGGAAGGAGGCGGTAAGGGGAGTGCGAGAGAAGAACTT 60
Db 409 GGAGCGCGAGGTCGGGATGATCCTGGAAGGAGGCGGTAAGGGGAGTGCGAGAGAAGAA 468
QY 61 CTGGAATAATGGCAAAAAAAGTAAAAAAAATGAGAAGAAAGAAAGAAACCAACTGTGAG 120
Db 469 CTTTITTTAACTGAACATAAAGTGAAAAAGATAAGAAAGAAAGAAACCAACTGTGAG 528
QY 121 CAGCTTTGCAATGTTTCCGCTATTAATGCTTCATAGTGTGTATATGTTGGTGGGAC 180
Db 529 TGTATTTCAATGTTTCCGCTATTAATGCTTCATAGTGTGTATATGTTGGTGGGAC 588
QY 181 ATGCTGCGCATCATCCATGGAGCTGCACCTCCCTCATGATGCTGTTTGGGAAACAT 240
Db 589 TTTGGCTGCGCATCATCCATGGGCTGGAGCTTCCCTCATGATGCTGTTGTTGGGAAAT 648
QY 241 GACAGATAGCTTTGCAATGCAAGTAATTTCAAGAAACAAAACCTTTCCAGTGTATAATTA 300
Db 649 GACAGATATCTTTGCAATGCAAGTAATTTCAAGATCTGATGT-----CAAACATCAC 702
QY 301 TGAAGATATTAGCAACAATACACACATTTTCATCAACCATCTGGAGGAGGAAATGCCAC 360
Db 703 TAATAGAGTATATCAATGATACAGGGTTCCTTCAATGAATCTGGAGGAGGATGACCCAG 762
```


Db 2923 TGCTGTAATACCCAGAAATATAGCAATCTTGGACAGGAATAATATATATCCCTCATCTA 2982
Qy 2581 TGGTTGGCAATTAACACTTTTACTCTTAGCAATTTGTACCACCATCATTTGCAATAGCAGGAGT 2640
Db 2983 TGGTTGGCAACTAACACTTGTACTCTTAGCAATTTGTACCACCATCATTTGCAATAGCAGGAGT 3042
Qy 2641 TGTGAAATGAAATGTTGTCTGGACAAAGCACTGAAAGATAAGAAAGAGCTAGAAAGGAGC 2700
Db 3043 TGTGAAATGAAATGTTGTCTGGACAAAGCACTGAAAGATAAGAAAGAGCTAGAAAGGAGC 3102
Qy 2701 TGGGAAGATGCTACAGAAGCCATCGAANAACCTCCGAACTGTTGTTCTTTGACTCGGGA 2760
Db 3103 TGGGAAGATGCTACAGAAGCAATAGAAACCTCCGAACTGTTGTTCTTTGACTCGGGA 3162
Qy 2761 GCAGAAGTTTGAATACATGTATGCACAGAGTTTGCAGTACCAATACAGAAACCTTTTGAG 2820
Db 3163 GCAGAAGTTTGAACATATGTATGCTCAGAGTTTGCAGTACCAATACAGAAACCTTTTGAG 3222
Qy 2821 GAAAGCACACATCTTCGGGGTCTCATTTTCTATCACCCAGCAATGATGATTTTTCCTA 2880
Db 3223 GAAAGCACACATCTTTGGAATTTACATTTTCTTCCACCCAGCAATGATGATTTTTCCTA 3282
Qy 2881 TGTGCTGCTTTCCGGTTTGGTGCTACTTGTGGCAAAATGAGTTCATGAACTTTTCAGGA 2940
Db 3283 TGTGCTGATTTTCCGGTTTGGAGCCTACTTGTGGCACATAAACAATATGAGCTTTGAGGA 3342
Qy 2941 TGTCTTTTGGTATCTCAGCTATTGTCTTTGGTGGCAATGAGTTCAGTTCAGTTC 3000
Db 3343 TGTCTTGTAGTATTTCAGCTTGTCTTTGGTGCCATGCGCTGGGGCAAGTGAGTTC 3402
Qy 3001 ATTTGCTCTGACTATGCCAAAGCCAAAGTATACAGAGCCACGCTCATCATGATCATTTGA 3060
Db 3403 ATTTGCTCTGACTATGCCAAAGCCAAAGTATACAGAGCCACATCATCATGATCATTTGA 3462
Qy 3061 AAAAAGCCCTCTGATTGACAGCTACAGCCCTCACGGCTCTCAAGCCAAATACGTTGGAGG 3120
Db 3463 AAAAAGCCCTTGTATTGACAGCTACAGCCGGAAGCCCTTAATGCCGACACATTTGGAAG 3522
Qy 3121 AAATGTCACATTTAATGAGTCTGTCTCAACTATCCACTCCAGCAGACATCCCGCTGCT 3180
Db 3523 AAATGTCACATTTTGGTGAAGTTGATTTCAACTATCCACTCCAGCAGACATCCCGCTGCT 3582
Qy 3181 CCAGGGCTGAGCTGAGCTGGAAGAGGCGCCAGAGCTGGCCCTGCTGTTAGTAGTCAAGTGG 3240
Db 3583 TCAGGGACTGAGCTGAGCTGGAAGAGGCGCCAGAGCTGGCTCTGTTGGGCGAGCAGTGG 3642
Qy 3241 CTGTGGGAAGAGCAGCTGTTGCTGAGCTCTAGAGCCCTTCTATGAGCCCTTGCTGGTTC 3300
Db 3643 CTGTGGGAAGAGCAGCTGTTGCTGAGCTCTAGAGCCCTTCTAGAGCCCTTGCTGGAGGNA 3702
Qy 3301 AGTCTAATTCATGCAAGAGATAAAGCACCTGAATGTCTGAGTGGCTCCGAGCACACCT 3360
Db 3703 AGTCTGCTGTTGATGGCAAGAAATAAAGCACCTGAATGTCTGAGTGGCTCCGAGCACACCT 3762
Qy 3361 GGGCATCTGCTCAGGAGCCCATCTCTGTTGACTGAGCATTTGCCGAGAACATTTGCCCTA 3420
Db 3763 GGGCATCTGCTCCAGGAGCCCATCTCTGTTGACTGAGCATTTGCCGAGAACATTTGCCCTA 3822
Qy 3421 TGGAGACAACAGCCGGTCTGATACATGAGAGATTTATCGAGGAGCCCAAGAGGCCAA 3480
Db 3823 TGGACACAACAGCCGGTGTGTACAGGAAGAGATCGTAGGGCAGCAAGAGGCCAA 3882
Qy 3481 CATACACACCTTCATCGAGACACTCCCTGTGAGAAATACAAACACAGAGTAGGAGCAAGG 3540
Db 3883 CATACATGCTTCATCGAGTCACTGCTCTAATAATATAGCACTAAAGTAGGAGCAAGG 3942
Qy 3541 AAACAGCTCTCTGTTGGCCAGAAACAGCGCATTTGCCATAGCTCGCCCTCTGTTAGACA 3600
Db 3943 AACTCAGCTCTCTGTTGGCCAGAAACAGCGCATTTGCCATAGCTCGTCCCTTGTGAGACA 4002
Qy 3601 GCCTCATATTTTCTTTTGGATGAAGCTTACATCAGCTCTGGATACAGAAAGTGAAGAGGT 3660
Db 4003 GCCTCATATTTTCTTTTGGATGAAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGAGGT 4062

Qy 3661 TGTCCAAGAGCCCTTGACAAAGCCAGAGAGGCCGACCTGCAATTTGTGATGCCACCG 3720
Db 4063 TGTCCAAGAGCCCTTGACAAAGCCAGAGAGGCCGACCTGCAATTTGTGATTTGCTCACCG 4122
Qy 3721 CTGTGCCACCATCCAGAAATGAGTAAATAGTGGTGTTCAGAAATGGCAAGTCAAGGA 3780
Db 4123 CTGTGCCACCATCCAGAAATGAGTAAATAGTGGTGTTCAGAAATGGCAAGTCAAGGA 4182
Qy 3781 GCATGGCACACATCAACAGCTGCTGGCCAGAAAGGATCATTTTTCATGCTCAGTGT 3840
Db 4183 GCATGGCACAGCATCAGAGCTGCTGGCCAGAAAGGATCATTTTTCATGCTCAGTGT 4242
Qy 3841 CCAGCTCGAGCAAGCGCTAGTGAACCTGTGGCCATATGAGCTGTGTAATATTTTAAAT 3900
Db 4243 CCAGCTCGAACAAGCGCCAGTGAACCTCTGACTGTATGAGATGTTAAATACATTTTAAAT 4302
Qy 3901 ATTTGTTTAAACATGCAATTTAATCAAACTTAAAGGTTGAGCCTTACTGGAAAAAAT 3960
Db 4303 ATTT--TGTTAGATATGACATTTAATCAAACTTAAAGTGTGAGCCTTACTGGAAAAAT 4355
Qy 3961 ATGTAGAACTACTCTTTAAACATTTCTTGTGCAACTGAAGATCATTTCCACCAAGTTCAG 4020
Db 4356 ATGAAGAGTACTCTTTAAACATTTCTTCACTCACTCACTTCAGAGTCTTC----- 4403
Qy 4021 AGTCTTCAGATTTTATATTAAGGAACCA--RAGAAACATTTATCTGATGGAATAAAAT 4078
Db 4404 ----AGAGACTTCGTAATTAAGGAACAGAGTGAGACATCATCAAGTGGAGAGAAAT 4458
Qy 4079 ACTGGTGTAAATGCAATATAAAATATAGAGTAAATCAAACTAGATTTTGTAAATAAT 4138
Db 4459 CATAGTGTAAACGCTATATAATTTTATACAGAAATTAAGTAGATTTTAAAGATAAA 4518
Qy 4139 TGTATAATTTTGTATATATTTT-----ATTTGTACTTACTGCTTCTGAAAGAT 4190
Db 4519 ATGTGTAATTTTGTATATTTTCCCATTTGGACTGTAACTGACGCTTCTGCTAAAGAT 4578
Qy 4191 TATAGAGTGGTAAAGTACTG--AATGTTTGAATAAAGTGTAGCTATAATAAACTAA 4249
Db 4579 TATAGAGTAGCAAAAGTATTTGAAATGTTTGCATAAAGTG---TCTATAATAAACTAA 4635
Qy 4250 ACTTTTATATCAAAAAAATAAAAAA 4279
Db 4636 ACTTCATGTGAAAAAATAAAAAA 4665

RESULT 5
US-08-784-649A-1
; Sequence 1, Application US/08784649A
; Patent No. 5830697
; GENERAL INFORMATION:
; APPLICANT: Sikic, Branimir I
; APPLICANT: Chen, Gang
; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
; TITLE OF INVENTION: CYCLOSPORIN MODULATION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/784,649A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

QY 1921 TGAATTCAGTTAGAAAATGCCACTGGTGAATCCAAAAGTGAAGTGATGCCCTTGGAAAT 1980
Db 1922 TGAATTCAGTTAGAAAATGCCACTGGTGAATCCAAAAGTGAAGTGATGCCCTTGGAAAT 1981
QY 2037 TGAAGTTGAATAGAAAATGACGTGATGAATCCAAAAGTGAATTCGATGCCCTTGGAAAT 2096
Db 2038 TGAAGTTGAATAGAAAATGACGTGATGAATCCAAAAGTGAATTCGATGCCCTTGGAAAT 2097
QY 1981 GTCTCCAAAAGATTACAGGGTCCAGTTTAAATAAAAGAGATCAACTCGCAGGAGTATACA 2040
Db 1982 GTCTCCAAAAGATTACAGGGTCCAGTTTAAATAAAAGAGATCAACTCGCAGGAGTATACA 2041
QY 2097 GTCTCCAAAAGATTACAGATCCAGTCTAATAAGAAAAGATCAACTCGTAGGAGTGTCCG 2156
Db 2098 GTCTCCAAAAGATTACAGATCCAGTCTAATAAGAAAAGATCAACTCGTAGGAGTGTCCG 2157
QY 2041 TGCACACAGGCCCAAGACAGAAAAGCTTGGTACAAAAGAGGACTTGAATGAGAATGTACC 2100
Db 2042 TGCACACAGGCCCAAGACAGAAAAGCTTGGTACAAAAGAGGACTTGAATGAGAATGTACC 2101
QY 2157 TGGATCAAGGCCCAAGACAGAAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGTATACC 2216
Db 2158 TGGATCAAGGCCCAAGACAGAAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGTATACC 2217
QY 2101 TCCAGTTTCTCTCTGGAGGATTTCTGAAGCTGAACCTCAACTGAATGGGCTTAATTTTGGT 2160
Db 2102 TCCAGTTTCTCTCTGGAGGATTTCTGAAGCTGAACCTCAACTGAATGGGCTTAATTTTGGT 2161
QY 2217 TCCAGTTTCTCTCTGGAGGATTTATGAAGCTAAATTTAACTGAATGGGCTTAATTTTGGT 2276
Db 2218 TCCAGTTTCTCTCTGGAGGATTTATGAAGCTAAATTTAACTGAATGGGCTTAATTTTGGT 2277
QY 2161 TGGTATATTTTGTGCTATTAATAACGGGAGGCTGCAACCCAGCATTTCAATAATATTTTC 2220
Db 2162 TGGTATATTTTGTGCTATTAATAACGGGAGGCTGCAACCCAGCATTTCAATAATATTTTC 2221
QY 2277 TGGTATATTTTGTGCTATTAATAATGGAGGCTGCAACCCAGCATTTGCAATAATATTTTC 2336
Db 2278 TGGTATATTTTGTGCTATTAATAATGGAGGCTGCAACCCAGCATTTGCAATAATATTTTC 2337
QY 2221 AAGGATATAGGATCTTTACCGGAGATGAGGATCCTGAAAACAAAACACAGAAATAGTAA 2280
Db 2222 AAGGATATAGGATCTTTTACAGAAATGATGATCCTGAAACAAAACACAGAAATAGTAA 2281
QY 2337 AAGGATATAGGAGTTTTCAGAAATGATGATCCTGAAACAAAACACAGAAATAGTAA 2396
Db 2338 AAGGATATAGGAGTTTTCAGAAATGATGATCCTGAAACAAAACACAGAAATAGTAA 2397
QY 2281 CATGTTTCTCTGATTTCTTAGTCCTTGGAAATATATTTCTTTTATTAATATTTTCCCTCA 2340
Db 2282 CATGTTTCTCTGATTTCTTAGTCCTTGGAAATATATTTCTTTTATTAATATTTTCCCTCA 2341
QY 2397 CTGTTTTCACATTTCTTAGCTAGCCCTTGGAAATATATTTCTTTTATTAATATTTTCCCTCA 2456
Db 2398 CTGTTTTCACATTTCTTAGCTAGCCCTTGGAAATATATTTCTTTTATTAATATTTTCCCTCA 2457
QY 2341 GGGCTTCACATTTGGCANAGCTGGGAGATCCTCACTAAGCGCTTCGATACATGTTT 2400
Db 2342 GGGCTTCACATTTGGCANAGCTGGGAGATCCTCACTAAGCGCTTCGATACATGTTT 2401
QY 2457 GGGTTTCACATTTGGCANAGCTGGAGATCTTCACCAAGCGCTCCGATACATGTTT 2516
Db 2458 GGGTTTCACATTTGGCANAGCTGGAGATCTTCACCAAGCGCTCCGATACATGTTT 2517
QY 2401 CAGATCCATGCTGAGACAGGATCTAGCTGGTTGATGACCCCTAAAACACCACCTGGAGC 2460
Db 2402 CAGATCCATGCTGAGACAGGATCTAGCTGGTTGATGACCCCTAAAACACCACCTGGAGC 2461
QY 2517 CCGATCCATGCTCAGACAGGATGTAGTTGGTTGATGACCCCTAAAACACCACCTGGAGC 2576
Db 2518 CCGATCCATGCTCAGACAGGATGTAGTTGGTTGATGACCCCTAAAACACCACCTGGAGC 2577
QY 2461 ATTGACAAACAGGCTTCCCAATGATCGGCTCAAGTTAAAGGGCTATAGTTTCCAGGCT 2520
Db 2462 ATTGACAAACAGGCTTCCCAATGATCGGCTCAAGTTAAAGGGCTATAGTTTCCAGGCT 2521
QY 2577 ATTGACTACCAAGCTGCCAATGATGCTCAAGTTTAAAGGGCTATAGTTTCCAGGCT 2636
Db 2578 ATTGACTACCAAGCTGCCAATGATGCTCAAGTTTAAAGGGCTATAGTTTCCAGGCT 2637
QY 2521 TGCCTCATTTACCCAGAAATATAGCAATCTTTGGACAGGATATATATATATATATATATAT 2580
Db 2522 TGCCTCATTTACCCAGAAATATAGCAATCTTTGGACAGGATATATATATATATATATATAT 2581
QY 2637 TGCCTCATTTACCCAGAAATATAGCAATCTTTGGACAGGATATATATATATATATATATAT 2696
Db 2638 TGCCTCATTTACCCAGAAATATAGCAATCTTTGGACAGGATATATATATATATATATATATAT 2697
QY 2581 TGGTTGGCAATTAACATTTTACTTTAGCAATTTGACCAATCTTGCATATAGCAGGAGT 2640
Db 2582 TGGTTGGCAATTAACATTTTACTTTAGCAATTTGACCAATCTTGCATATAGCAGGAGT 2641
QY 2697 TGGTTGGCAATTAACATTTTACTTTAGCAATTTGACCAATCTTGCATATAGCAGGAGT 2756
Db 2698 TGGTTGGCAATTAACATTTTACTTTAGCAATTTGACCAATCTTGCATATAGCAGGAGT 2757
QY 2641 TGTGAAATGAAATGTTGCTGGACAGCACCTGAAAGATAAGAAAGAGCTAGAAGGAGC 2700
Db 2642 TGTGAAATGAAATGTTGCTGGACAGCACCTGAAAGATAAGAAAGAGCTAGAAGGAGC 2701
QY 2757 TGTGAAATGAAATGTTGCTGGACAGCACCTGAAAGATAAGAAAGAGCTAGAAGGAGTGC 2816
Db 2758 TGTGAAATGAAATGTTGCTGGACAGCACCTGAAAGATAAGAAAGAGCTAGAAGGAGTGC 2817
QY 2701 TGGGAAGATTGCTACAGAAGCCATCGAAAACCTTCCGAACCTGTTGTTCTTTGACTCGGGA 2760
Db 2702 TGGGAAGATTGCTACAGAAGCCATCGAAAACCTTCCGAACCTGTTGTTCTTTGACTCGGGA 2761
QY 2817 TGGGAAGATTGCTACTGAAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTGACTCAGGA 2876
Db 2818 TGGGAAGATTGCTACTGAAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTGACTCAGGA 2877
QY 2761 GCAGAAGTTTGAATACATGATGATGCTCAGAGTTTGCAGGTACCAATACAGAACTCTTTGAG 2820
Db 2762 GCAGAAGTTTGAATACATGATGATGCTCAGAGTTTGCAGGTACCAATACAGAACTCTTTGAG 2821
QY 2877 GCAGAAGTTTGAATACATGATGATGCTCAGAGTTTGCAGGTACCAATACAGAACTCTTTGAG 2936
Db 2878 GCAGAAGTTTGAATACATGATGATGCTCAGAGTTTGCAGGTACCAATACAGAACTCTTTGAG 2937
QY 2821 GAAAGCACACATCTTTCCGGGGTCTCATTTTCTATCACCAGCAATGATGATTTTTCCTTA 2880
Db 2822 GAAAGCACACATCTTTCCGGGGTCTCATTTTCTATCACCAGCAATGATGATTTTTCCTTA 2881
QY 2937 GAAAGCACACATCTTTGGAATTAACATTTTCTTCCAGGCAATGATGATTTTTCCTTA 2996
Db 2938 GAAAGCACACATCTTTGGAATTAACATTTTCTTCCAGGCAATGATGATTTTTCCTTA 2997
QY 2881 TCGTGGCTTTTCCGGTTTGGTGCTACTTGGTGGCAATGAGTTTCAATGATTTTTCAGGA 2940
Db 2882 TCGTGGCTTTTCCGGTTTGGTGCTACTTGGTGGCAATGAGTTTCAATGATTTTTCAGGA 2941
QY 2997 TCGTGGATTTTCCGGTTTGGGAGCTACTTGGTGGCAATTAACATCAAGCTTTGAGGA 3056
Db 2998 TCGTGGATTTTCCGGTTTGGGAGCTACTTGGTGGCAATTAACATCAAGCTTTGAGGA 3057
QY 2941 TGTCTTTTGGTATTTCTCAGCTATTTGCTTTGGTGCCATGCGAGTGGGCGAGGTCAGTTC 3000
Db 2942 TGTCTTTTGGTATTTCTCAGCTATTTGCTTTGGTGCCATGCGAGTGGGCGAGGTCAGTTC 3001
QY 3057 TGTCTCTGTTAGTATTTTTCAGCTGTTGCTTTGGTGCCATGCGCGTGGGCGAAGTCAGTTC 3116
Db 3058 TGTCTCTGTTAGTATTTTTCAGCTGTTGCTTTGGTGCCATGCGCGTGGGCGAAGTCAGTTC 3117

QY 3001 ATTTGCTCTGACTATGCCAAAGCCAAAGATGATCAGAGGCCACGTCATCATGATCATTCGA 3060
Db 3002 ATTTGCTCTGACTATGCCAAAGCCAAAGATGATCAGAGGCCACGTCATCATGATCATTCGA 3061
QY 3117 ATTTGCTCTGACTATGCCAAAGCCAAATATCAGCAGCCACATCATGATCATTCGA 3176
Db 3118 ATTTGCTCTGACTATGCCAAAGCCAAATATCAGCAGCCACATCATGATCATTCGA 3177
QY 3061 AAAAAGCCCTCTGATTTGACAGCTACAGCCCTCAGGGCTCAAGCCAAATACGTTGGAAGG 3120
Db 3062 AAAAAGCCCTCTGATTTGACAGCTACAGCCCTCAGGGCTCAAGCCAAATACGTTGGAAGG 3121
QY 3177 AAAAAGCCCTTTGATTTGACAGCTACAGCAGGAAGGCTTAATGCCGNACACATTTGGAAGG 3236
Db 3178 AAAAAGCCCTTTGATTTGACAGCTACAGCAGGAAGGCTTAATGCCGNACACATTTGGAAGG 3237
QY 3121 AAATGTGACATTTAATGAGTCTGTTCAACTATCCCACTCGACACACATCCCGTGTCT 3180
Db 3122 AAATGTGACATTTGATGAGTCTGTTCAACTATCCCACTCGACACACATCCCGTGTCT 3181
QY 3237 AAATGTGACATTTGATGAGTCTGTTCAACTATCCCACTCGACACACATCCCGTGTCT 3296
Db 3238 AAATGTGACATTTGATGAGTCTGTTCAACTATCCCACTCGACACACATCCCGTGTCT 3297
QY 3181 CCAGGGCTGAGCCCTCGAGGTGAAGAGGGCCAGACGCTGCCCTCTGATGAGCAGTGG 3240
Db 3182 CCAGGGCTGAGCCCTCGAGGTGAAGAGGGCCAGACGCTGCCCTCTGATGAGCAGTGG 3241
QY 3297 TCAGGAGCTGAGCCCTGGAGGTGAAGAGGGCCAGACGCTGCCCTCTGATGAGCAGTGG 3356
Db 3298 TCAGGAGCTGAGCCCTGGAGGTGAAGAGGGCCAGACGCTGCCCTCTGATGAGCAGTGG 3357
QY 3241 CTGTGGGAAGACACAGTTGTTCAAGTCTAGAGGCTTCTATGACCCCTTGGCTGGTTC 3300
Db 3242 CTGTGGGAAGACACAGTTGTTCAAGTCTAGAGGCTTCTATGACCCCTTGGCTGGTTC 3301
QY 3357 CTGTGGGAAGACACAGTGGTCCAGCTCCTGGAGGGTTCACGACCCCTTGGCAGGAA 3416
Db 3358 CTGTGGGAAGACACAGTGGTCCAGCTCCTGGAGGGTTCACGACCCCTTGGCAGGAA 3417
QY 3301 AGTGTCTAATTTGATGCAAGAGATAAAGCAGCTGAATGCCAGTGGCTCCGAGCACACCT 3360
Db 3302 AGTGTCTAATTTGATGCAAGAGATAAAGCAGCTGAATGCCAGTGGCTCCGAGCACACCT 3361
QY 3417 AGTGTCTGTTGATGGCAAGAAATAAAGCAGCTGAATTTTCAGTGGCTCCGAGCACACCT 3476
Db 3418 AGTGTCTGTTGATGGCAAGAAATAAAGCAGCTGAATTTTCAGTGGCTCCGAGCACACCT 3477
QY 3361 GGGCATGCTCTCAGGAGCCCATCCTGTTGACGTGACGATTTGCCGAGAACATTTGCCCTA 3420
Db 3362 GGGCATGCTCTCAGGAGCCCATCCTGTTGACGTGACGATTTGCCGAGAACATTTGCCCTA 3421
QY 3477 GGGCATGCTCTCAGGAGCCCATCCTGTTGACGTGACGATTTGCCGAGAACATTTGCCCTA 3536
Db 3478 GGGCATGCTCTCAGGAGCCCATCCTGTTGACGTGACGATTTGCCGAGAACATTTGCCCTA 3537
QY 3421 TGGAGACAAACGCGGCTGATACATGAAGAGATTTATGAGGCGACGCAAGAGGCGCAA 3480
Db 3422 TGGAGACAAACGCGGCTGATACATGAAGAGATTTATGAGGCGACGCAAGAGGCGCAA 3481
QY 3537 TGGAGACAAACGCGGCTGATGTCAGAGAGAGATTTGTGAGGCGACGCAAGAGGCGCAA 3596
Db 3538 TGGAGACAAACGCGGCTGATGTCAGAGAGAGATTTGTGAGGCGACGCAAGAGGCGCAA 3597
QY 3481 CATACACCACTTCAATCGAGACACTCCTGTGAGAAATACAAACCCAGAGTAGGAGCAAGG 3540
Db 3482 CATACACCACTTCAATCGAGACACTCCTGTGAGAAATACAAACCCAGAGTAGGAGCAAGG 3541
QY 3597 CATACATGCTTCAATCGAGTCACTGCTTAATAATAATATAGCACTAAAGTAGGAGCAAGG 3656
Db 3598 CATACATGCTTCAATCGAGTCACTGCTTAATAATAATATAGCACTAAAGTAGGAGCAAGG 3657
QY 3541 AACCCAGCTCTCTGSGGCGCAGAAACAGCGCATTTGCCATAGCTCCGCTCTTTTAGACA 3600
Db 3542 AACCCAGCTCTCTGSGGCGCAGAAACAGCGCATTTGCCATAGCTCCGCTCTTTTAGACA 3601
QY 3657 AACTGAGCTCTCTGSGGCGCAGAAACAGCGCATTTGCCATAGCTCCGCTCTTTTAGACA 3716
Db 3658 AACTGAGCTCTCTGSGGCGCAGAAACAGCGCATTTGCCATAGCTCCGCTCTTTTAGACA 3717
QY 3601 GCCTCATATTTTCTTTGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGT 3660
Db 3602 GCCTCATATTTTCTTTGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGT 3661
QY 3717 GCCTCATATTTTCTTTGATGAAGCCAGCTCAGCTCTGATACAGAAAGTGAAGAGT 3776
Db 3718 GCCTCATATTTTCTTTGATGAAGCCAGCTCAGCTCTGATACAGAAAGTGAAGAGT 3777
QY 3661 TGTCCAAAGAGCCCTGGCAAAAGCCAGAGAGGCGCACCTGATTCGCCCAACG 3720
Db 3662 TGTCCAAAGAGCCCTGGCAAAAGCCAGAGAGGCGCACCTGATTCGCCCAACG 3721
QY 3777 TGTCCAAAGAGCCCTGGCAAAAGCCAGAGAGGCGCACCTGATTCGCCCAACG 3836
Db 3778 TGTCCAAAGAGCCCTGGCAAAAGCCAGAGAGGCGCACCTGATTCGCCCAACG 3837
QY 3721 CTTGTCCACCATCCAGAAATGCAGATTTAATAGTGGTGTTCAGAAATGGAAGTCAAGGA 3780
Db 3722 CTTGTCCACCATCCAGAAATGCAGATTTAATAGTGGTGTTCAGAAATGGAAGTCAAGGA 3781
QY 3837 CCTGTCCACCATCCAGAAATGCAGACTTAATAGTGGTGTTCAGAAATGGAAGTCAAGGA 3896
Db 3838 CCTGTCCACCATCCAGAAATGCAGACTTAATAGTGGTGTTCAGAAATGGAAGTCAAGGA 3897
QY 3781 GCATGGCACACATCAACAGCTGCTGCCCAAGAAAGGCACTATTTTCCATGCTCAGTGT 3840
Db 3782 GCATGGCACACATCAACAGCTGCTGCCCAAGAAAGGCACTATTTTCCATGCTCAGTGT 3841
QY 3897 GCATGGCACGCTCAGCAGCTGCTGCCCAAGAAAGGCACTATTTTCCATGCTCAGTGT 3956
Db 3898 GCATGGCACGCTCAGCAGCTGCTGCCCAAGAAAGGCACTATTTTCCATGCTCAGTGT 3957
QY 3841 CCAGCTGGAGCAAGCGCTAGTGAACCTGTGGCCATATGAGCTGTGTTAAATATTTTAAAT 3900
Db 3842 CCAGCTGGAGCAAGCGCTAGTGAACCTGTGGCCATATGAGCTGTGTTAAATATTTTAAAT 3901
QY 3957 CCAGCTGGAGCAAGCGCCAGCTGACTCTGATGATGAGTGTAAATATCTTTTAAAT 4016
Db 3958 CCAGCTGGAGCAAGCGCCAGCTGACTCTGATGATGAGTGTAAATATCTTTTAAAT 4017
QY 3901 ATTTTGTGTTAAACATGGCATTTAATCAAAGTTTAAAGGTGAGCACTTACTGGAAAACT 3960
Db 3902 ATTTTGTGTTAAACATGGCATTTAATCAAAGTTTAAAGGTGAGCACTTACTGGAAAACT 3961
QY 4017 ATT--TGTGTAGATATGACATTTATTCAAAGTTAAA-----AGCAACACTTACAGAAAT 4069
Db 4018 ATT--TGTGTAGATATGACATTTATTCAAAGTTAAA-----AGCAACACTTACAGAAAT 4069
QY 3961 ATGTAGAACTACCTGTTTAAACATTTCTTGTGCAACTGGAAGATCATTCACCAAGTTCAG 4020
Db 3962 ATGTAGAACTACCTGTTTAAACATTTCTTGTGCAACTGGAAGATCATTCACCAAGTTCAG 4021
QY 4070 ATGGAAGGATATCTGTTTAAACATTTCTCAGTCAAGTTTCAGAGTCTC-----4117
Db 4071 ATGGAAGGATATCTGTTTAAACATTTCTCAGTCAAGTTTCAGAGTCTC-----4118
QY 4021 AGCTTTCAGATTTTATATTAAGGAAACCAA--AAGAAACATTTCTGATGGAATAAAT 4078
Db 4022 AGCTTTCAGATTTTATATTAAGGAAACCAA--AAGAAACATTTCTGATGGAATAAAT 4079
QY 4118 -----AGAGACTTCGTAATTAAGGAAACAGAGTGAAGACATCATCAAGTGGAGAGAAAT 4172
Db 4119 -----AGAGACTTCGTAATTAAGGAAACAGAGTGAAGACATCATCAAGTGGAGAGAAAT 4173
QY 4079 ACTGCTGTTAATGCAATTTAAATATAGAGTAAATTCAAAAGTAGATTTTGTAAATAAAT 4138
Db 4080 ACTGCTGTTAATGCAATTTAAATATAGAGTAAATTCAAAAGTAGATTTTGTAAATAAAT 4139

Db 4173 CATAGTTTAACTGCATTATAAATTTTATACAGAAATAAAGTAGATTTAAGATAAA 4232
QY 4139 TGTATAATTTTGGTTTATATTTT 4161
Db 4233 ATCTGTAATTTTGGTTTATATTTT 4255

RESULT 6

```

US-08-784-649A-5
; Sequence 5, Application US/08784649A
; Patent No. 5830697
; GENERAL INFORMATION:
; APPLICANT: Sikic, Branimir I
; APPLICANT: Chen, Gang
; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
; TITLE OF INVENTION: CYCLOSPORIN MODULATION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08784,649A

```

Query Match 76.4%; Score 3270.6; DB 2; Length 4264;
Best Local Similarity 87.9%; Pred. No. 0;
Matches 3658; Conservative 0; Mismatches 469; Indels 36; Gaps 7;

QY	301	TGAAAGTATTACGAACAATAACACACATTTTCATCAACCATCTCGAGAGGAGAAATGACCAC	360
Db	420	TAATGAAGTGATATCAATGATACAGGGTCTTTCATGAATCTGGAGGAACACATGACCAG	479
QY	361	GTATGCTTATTATACAGTGGGATCGGTGCGTCTGGTGGCTGCTTTACATCAGGT	420
Db	480	GTATGCTTATTATACAGTGGAAATGGTGGCTGGGGTGCTGGTGGCTGCTTTACATTCAGGT	539
QY	421	TTCAATTCGGTGCCTGGCAGCAGGAACAGATACTCAAAATTAGAAAACAAATTTTTC	480
Db	540	TTCAATTTGGTGCCTGGCAGCTGGAAGACAAATACACAAAATAGAAACAGTTTTC	599
QY	481	TGCTATCATGCGACAGGAGATGGCTGGTTGACGTGCATGACGTGGGGAGCTTAAAC	540
Db	600	TGCTATTAATGGCAGCAGGAGATAGGCTGGTTGATGTGCACGATGTTGGGAGCTTAAAC	659
QY	541	CGGGCTCACAGCAGATGTCTCCAAAATCAATGAAGGAATTTGGCGACAAAATTTGGAATGT	600
Db	660	CGGACTTACAGATGATGTCTCCAAGATTATGAAGGAATTTGGTGACAAAATTTGGAATGT	719
QY	601	CTTTTCACTCAATAGCAACATTTTTCACGGGTTTTATAGTGGGTTTTACACGTGGTTGAA	660
Db	720	C---CAGTCAATGGCAACATTTTTCACGTGGGTTTTATAGTAGGATTTACACGTGGTTGAA	776
QY	661	GCTAACCCCTGTGATTTTGGGCATCAGCCCTGTCTTGGACTTTCAGCGCCCATCTGGC	720
Db	777	GCTAACCCCTGTGATTTTGGCCATCAGTCTCTGTCTTGGACTGTCACTGTCTTGGGC	836
QY	721	AAAGATACTATCTCAATTTACTGATAAAGAACTCTTGGCTATGCAAAAAGCTGGAGCAGT	780
Db	837	AAAGATACTATCTCAATTTACTGATAAAGAACTCTTAGCGTATGCAAAAAGCTGGAGCAGT	896
QY	781	AGCTGAAGAAGCTTTAGCAGCAATCAGAACTGTGATTTGGCTTTGGAGGACAAAAGAA	840
Db	897	AGCTGAAGAGCTCTTGGCAGCAATTAGACTGTGATTTGGATTTGGAGGACAAAAGAA	956
QY	841	ACTTGAAGGTACACAAAAATTTGAAGAAGCTAAAGGAATTTGGGATTAAGAAAGCTAT	900
Db	957	ACTTGAAGGTACACAAAAATTTGAAGAAGCTAAAGGAATTTGGGATTAAGAAAGCTAT	1016
QY	901	CACGGCCAAATTTCTATTTGGTGCCTTCTTATTTGATCTATGATCATATGCTCTGGC	960
Db	1017	TACAGCCAAATTTCTATAGGTGCTGCTTTCTGCTGTGATCTATGATCTTATGCTCTGGC	1076
QY	961	TTCTGTGATGGGACCTCTTGGTCTCTCCAGTGAATATCTATTTGACAAAGTACTCAC	1020
Db	1077	CTTCTGTGATGGGACCACTTGGTCTCTCAGGGGAATTTCTATTTGGACAAAGTACTCAC	1136
QY	1021	TGCTCTCTTTCTGTATTAATTTGGGCTTTTATTTGACAGGCAATCCCAAGCATTTGA	1080
Db	1137	TGTATTTCTTTCTGTATTAATTTGGGCTTTTATTTGGGCTTTTATTTGACAGGCAATTTGA	1196
QY	1081	AGCAATTTGCAAGCGAAGAGGACGCTTATGAAATCTTTCAAGATAATTTGACAAATAACC	1140
Db	1197	AGCAATTTGCAAAATGCAAGAGGACGCTTATGAAATCTTTCAAGATAATTTGATAAAGCC	1256
QY	1141	AAGCATTGACAGCTATTCGAAGAGTGGACATAAACAGATATAATTAAGGGAATTTGGA	1200
Db	1257	AAGTATTTGACAGCTATTCGAAGAGTGGGACAAAACAGATAATTAAGGGAATTTGGA	1316
QY	1201	ATTTCAAAATGTTCACTTCAGTTACCTTCTCGAAAAGCACTTAAGATCTTAAAGGCTCT	1260
Db	1317	ATTTCAAGAAATGTTCACTTCAGTTACCTTCTCGAAAAGCACTTAAGATCTTAAAGGCTCT	1376
QY	1261	CAACCTGAAGGTTACAGTGGGACAGACTGTGCGCTGGTTGGGAACAGTGGCTCGCGGAA	1320
Db	1377	GAACCTGAAGGTCAGAGTGGGACAGCGTGGCCCTGTTGGAAACAGTGGCTGGGA	1436
QY	1321	GACACGACCGTGCAGCTGATCCAGAGCTCTATGACCCACAGATGGCATGTCTGTAT	1380
Db	1437	GAGCAACAGTCCAGCTGATCAGAGGCTCTATGACCCACAGAGGGATGCTCAGTGT	1496

QY 1381 TGATGGACAGACATAGGACCATTAATGTAAAGCATCTTCGGAAATTAAGTGTGGT 1440
 Db 1497 TGATGGACAGATATAGGACCATTAATGTAAAGTCTTACGGGAATCAATGTGTGGT 1556
 QY 1441 GAGTCAGAGCCTGTGTTTGGCCACCACCATAGCTGAAACAAATTCGCTATGGCCGGA 1500
 Db 1557 GAGTCAGAGCCTGTGTTTGGCCACCACCATAGCTGAAACAAATTCGCTATGGCCGGA 1616
 QY 1501 AAATGTCACCATGGATGAGATTGAAAGCTGTTAAGGAAGCAATGCGCTATGATTTAT 1560
 Db 1617 AAATGTCACCATGGATGAGATTGAAAGCTGTTAAGGAAGCAATGCGCTATGATTTAT 1676
 QY 1561 CATGAAACTACCTAATAATTTGACACTCTGGTTGGAGAGAGAGGGCCAGCTGAGTGG 1620
 Db 1677 CATGAAACTGCTCATAAATTTGACACCCTGGTTGGAGAGAGAGGGCCAGTGTGAGTGG 1736
 QY 1621 TGGCAGAAACAGAGAAATCGCCATTTGCTGGGCCCTGGTTGCGAACCCCAAGATCTTCT 1680
 Db 1737 TGGCAGAAACAGAGAAATCGCCATTTGCGACGCTGGTTGCGAACCCCAAGATCTTCT 1796
 QY 1681 GCTGATGAGGCAACGTCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCT 1740
 Db 1797 GCTGATGAGGCAACGTCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCT 1856
 QY 1741 GGAATAGGCCAGAAAGGCGGACCTACCATTTGTATAGCTCATCTGTTGTCTACAGTTCCG 1800
 Db 1857 GGAATAGGCCAGAAAGGCGGACCTACCATTTGTATAGCTCATCTGTTGTCTACAGTTCCG 1916
 QY 1801 TAATGCGGATGTCATTCGTTGTTTGTATGATGAGTCAATTTGGAGAAAGGAATCATGA 1860
 Db 1917 TAATGCGGATGTCATTCGTTGTTTGTATGATGAGTCAATTTGGAGAAAGGAATCATGA 1976
 QY 1861 TGAACATGAAAGAGAGGGCAATTTACTTCAACTTGTCAAAATGTCAGACAGAGAGAAA 1920
 Db 1977 TGAACATGAAAGAGAGGGCAATTTACTTCAAACTTGTCAAAATGTCAGACAGAGAGAAA 2036
 QY 1921 TGAATTTGAGTTAGAAATGCCACTGGTGAATCCAAAGTGAAAGTGATGCGTTGGAAT 1980
 Db 2037 TGAATTTGAGTTAGAAATGCCACTGGTGAATCCAAAGTGAAATTTGATGCGTTGGAAT 2096
 QY 1981 GTCTCCAAAAGATTCAGGGTCCAGTTTAATAAAGAGATCAACTCCGACGAGGATATACA 2040
 Db 2097 GTCTCCAAAAGATTCAGGGTCCAGTTTAATAAAGAGATCAACTCCGACGAGGATATACA 2156
 QY 2041 TGCACCAAGGCCAAGACAGAAAGCTTGGTCAAAAGAGGACTTGAATGAGAATATACC 2100
 Db 2157 TGGATCAAGGCCAAGACAGAAAGCTTAGTACCAAGAGGCTCTGATGAAAGTATACC 2216
 QY 2101 TCCAGTTTCCTCTGGAGGATTCGAGCTGAACTCAACTGAATGGCCTTATTTGTGT 2160
 Db 2217 TCCAGTTTCCTCTGGAGGATTCGAGCTGAACTCAACTGAATGGCCTTATTTGTGT 2276
 QY 2161 TGGTATATTTTGTCTATTAACGAGGCGCTGCAACGACATTTTCAATAATATTTTC 2220
 Db 2277 TGGTATATTTTGTCTATTAACGAGGCGCTGCAACGACATTTTCAATAATATTTTC 2336
 QY 2221 AAGGATATAGGGATCTTTACCCGAGATGAGGATCCCTGAAACAAACAGCAGAAATAGTAA 2280
 Db 2337 AAGGATATAGGGATCTTTACCAAGAAATGATGATCCCTGAAACAAACAGCAGAAATAGTAA 2396
 QY 2281 CATGTTTTCGTATTTTCTAGTCCCTTGGAAATTTATTTTATACATTTTTCCTCCA 2340
 Db 2397 CTGTTTTCGTATTTTCTAGTCCCTTGGAAATTTATTTTATACATTTTTCCTCCA 2456
 QY 2341 GGGCTTCACATTTGGCAAGCTGGGGAGATCCTCAGTAAAGGCTTCGATACATGTTTTT 2400
 Db 2457 GGGTTTCACATTTGGCAAGCTGGAGAGATCCTCAGCAGGCTCCGATACATGTTTTT 2516
 QY 2401 CAGATCCATGCTGACAGAGATGTCAGTGGTTTGTATGACCCCTAATAACACCACTGAGC 2460
 Db 2517 CCGATCCATGCTGACAGAGATGTCAGTGGTTTGTATGACCCCTAATAACACCACTGAGC 2576
 QY 2461 ATTGACACACAGGCTTGCCAAATGATGGGCTCAAGTTAAAGGGGCTATAGTTCAGGCT 2520

Db 2577 ATTGACTACAGGCTCGCAATGATGCTCTCAAGTTAAAGGGGCTATAGTTCAGGCT 2636
 QY 2521 TCGTGTCAATCCCAAGATATAGCAATCTTGGACAGCAGCATTTATATATCTTATCTTA 2580
 Db 2637 TCGTGTAAATTTACCCAGATATAGCAATCTTGGACAGCAATTAATATATCTTATCTTA 2696
 QY 2581 TCGTGTGCAATTAACACTTTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGGAT 2640
 Db 2697 TCGTGTGCAATTAACACTTTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGGAT 2756
 QY 2641 TCGTGTGCAATTAACACTTTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGGAT 2700
 Db 2757 TCGTGTGCAATTAACACTTTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGGAT 2816
 QY 2701 TCGGAGATGCTACAGAGCCATCGAAACCTTCCGAACTTCTGTTTCTTTGACTCGGA 2760
 Db 2817 TCGGAGATGCTACAGAGCCATCGAAACCTTCCGAACTTCTGTTTCTTTGACTCAGGA 2876
 QY 2761 GCGAAGTTTGAATACATGATGACAGAGTTTGCAGAGTGGTGCAGTACAGAACTCTTTGAG 2820
 Db 2877 GCGAAGTTTGAATACATGATGACAGAGTTTGCAGAGTGGTGCAGTACAGAACTCTTTGAG 2936
 QY 2821 GAAAGCAGACATCTTCGGGCTCTCATTTCTATCACCCAGGCAATGATGATTTTCTTA 2880
 Db 2937 GAAAGCAGACATCTTCGGGCTCTCATTTCTATCACCCAGGCAATGATGATTTTCTTA 2996
 QY 2881 TGCTGGGCTGTTTCCGGTGTGGTGCCTACTTGGTGGCAAAATGAGTTTATGAACTTTCAGGA 2940
 Db 2997 TGCTGGGCTGTTTCCGGTGTGGTGCCTACTTGGTGGCAAAATGAGTTTATGAACTTTCAGGA 3056
 QY 2941 TGCTGCTGTTTGTATTCAGCTATTCAGCTATTCAGCTATTCAGCTATTCAGCTATTCAGCT 3116
 Db 3057 TGCTGCTGTTTGTATTCAGCTATTCAGCTATTCAGCTATTCAGCTATTCAGCTATTCAGCT 3176
 QY 3001 ATTTGCTCCTGACATGCAAGCAAGCAAAAGTATCAGAGCCCACTCATCATGATCATTTGA 3060
 Db 3117 ATTTGCTCCTGACATGCAAGCAAGCAAAATATCAGAGCCCACTCATCATGATCATTTGA 3176
 QY 3061 AAAAGCCCTCTGATGACAGCTACAGCCCTCAGGCTCAAGCCCAATAGTTTGGAGG 3120
 Db 3177 AAAAGCCCTCTGATGACAGCTACAGCCCTCAGGCTCAAGCCCAATAGTTTGGAGG 3236
 QY 3121 AAATGTCATTTAATGAGTGGTCTTCAACTATCCCACTCGACGACACATCCCGGCT 3180
 Db 3237 AAATGTCATTTGATGAGTGGTCTTCAACTATCCCACTCGACGACACATCCCGGCT 3296
 QY 3181 CCAGGGCTGAGCCTCGAGGTGAAGAGGGCCAGCGTGGCCCTCGTAGTACAGCTGG 3240
 Db 3297 TCAGGGACTGAGCCTCGAGGTGAAGAGGGCCAGCGTGGCTCTGTTGGCAGCAGTGG 3356
 QY 3241 CTGTTGGAGAGCAGATGTTTCAGCTCTTAGAGGCTTCTATGACCCCTTGGGCTGTC 3300
 Db 3357 CTGTTGGAGAGCAGATGTTTCAGCTCTTAGAGGCTTCTATGACCCCTTGGGCTGTC 3416
 QY 3301 AGTGTCAATTTGATGTCAGAGATTAAGCACTGAAATTCAGTGGCTTCCGAGCACCT 3360
 Db 3417 AGTGTCAATTTGATGTCAGAGATTAAGCACTGAAATTCAGTGGCTTCCGAGCACCT 3476
 QY 3361 GGGGATCGTGTCTAGGAGCCCATCTGTTGACCTGACAGATTCGCGAGAACATTCGCTA 3420
 Db 3477 GGGGATCGTGTCTAGGAGCCCATCTGTTGACCTGACAGATTCGCGAGAACATTCGCTA 3536
 QY 3421 TGGAGAACACAGCGGCTGCTATCATGAGAGATTTATGAGGAGCCCAAGAGGACCA 3480
 Db 3537 TGGAGAACACAGCGGCTGCTATCATGAGAGATTTATGAGGAGCCCAAGAGGACCA 3596
 QY 3481 CATACACCACTTCACTGAGACATCCCTGAGAAATACACACAGAGTAGGAGACAAAG 3540
 Db 3597 CATACACCTTCACTGAGACATCCCTGAGAAATACACACAGAGTAGGAGACAAAG 3656
 QY 3541 AAACCCAGCTCTGTTGGCCAGAACAGCGCATTTGCATAGCTCGGCTCTTTGTTAGACA 3600

```
Db 3657 AACTCAGCTCTCTGGTGGCCAGAAACACCGCATTCGCATAGCTCTGGCCCTTGTGTAGACA 3716
Qy 3601 CCCTCATATTTTGGTTTGGATGAAGTACATCAGCTCTGGATACAGAAAGTGAAGAAGT 3660
Db 3717 GCCTCATATTTTGGTTTGGATGAAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGAAGT 3776
Qy 3661 TGTCCAAAGAGCCCTTGACAAAGCCAGAGAGCCGACCTGCATTTGTATGATGCCCAAGG 3720
Db 3777 TGTCRAAGAGCCCTTGACAAAGCCAGAGAGCCGACCTGCATTTGTATGATGCCCAAGG 3836
Qy 3721 CTGTGTCACATCCAGATTCAGATTAATAGTGTGTTTTCAGAAATGGCAAGTCAAGGA 3780
Db 3837 CCTGTCCACCATCCAGATTCAGATTAATAGTGTGTTTTCAGAAATGGCAAGTCAAGGA 3896
Qy 3781 GCATGGCACATCAACAGCTCTGCGCCAGAGAGGCATCTATTTTCCATGGTCAGTGT 3840
Db 3897 GCATGGCACATCAACAGCTCTGCGCCAGAGAGGCATCTATTTTCCATGGTCAGTGT 3956
Qy 3841 CCAGGCTGGAGCAAGCGCTAGTGAAGCTGGCCATATGAGCTGTTAAATATTTTAAAT 3900
Db 3957 CCAGGCTGGAGCAAGCGCGAGTGAAGCTGTTAAATATTTTAAATATTTTAAAT 4016
Qy 3901 ATTTGTTTAAACATGGCAATTAATCAAAAGTTAAAGGTGAGCACTTACTCGAAAGT 3960
Db 4017 ATT--TGTTAGATATGACATTTATTCAAAGTTAAA-----AGCAAAACACTTACAGATT 4069
Qy 3961 ATGTAGAACTACTGTTTAAACATTTCTGCTGCAACTGAAGATCAATCCACCAAGTTCAG 4020
Db 4070 ATGAAGAGGTATCTGTTTAAACATTTCTGCTGCAAGTCAAGTTCAGAGTCTTC----- 4117
Qy 4021 AGTCTTCAGATTTTAAATTAAGGAACAA--AAGAAACATTTATCTGATGGAATAAAAT 4078
Db 4118 -----AGAGACTTCGTTAATTAAGCAACAGAGTGAAGACATCAAGTGGAGAGAAAT 4172
Qy 4079 ACTGGTGTAAATGCAATTAATAATTAAGTAAATTAAGTGAAGTAAATTTGTTTAAAT 4138
Db 4173 CATAGTTAAATGCAATTAATAATTTTAAACAGAAATTAAGTGAAGTAAATTTTAAAGATAA 4232
Qy 4139 TGTATAAATTTTGTATTTATTTT 4161
Db 4233 ATGTGTAATTTTGTATTTATTTT 4255
```

RESULT 7

```
US-08-793-610-5
; Sequence 5, Application US/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARBERS, Carol
; APPLICANT: OSTERTAG, Wolfram
; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
; TITLE OF INVENTION: FOR GENE TRANSFER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikolaïdo, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,610
; FILING DATE: 07-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 31 973.8
; FILING DATE: 08-SEP-1994
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: DE 195 03 952.1
; FILING DATE: 07-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/03175
; FILING DATE: 10-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berman, Richard J.
; REGISTRATION NUMBER: 39,105
; REFERENCE/DOCKET NUMBER: P1614-7007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6505 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
US-08-793-610-5
```

Query Match 75.7%; Score 3239.6; DB 2; Length 6505;

Best Local Similarity 89.6%; Pred. No. 0;

Matches 3495; Conservative 0; Mismatches 399; Indels 6; Gaps 1;

Qy 1 GGAGCCGAGGTCGGGATGGATCCCTGAAGGGCGCTAAGGGGAGTCGAGAGAAGAACTT 60

Db 1801 GGAGCCGAGGTCGGGATGGATCCCTGAAGGGCGCTAAGGGGAGTCGAGAGAAGAACTT 1860

Qy 61 CTGGAAATATGGCAAAATAAGTAAATAATGAGAAAGAAAGAAAGAAAGAAAGTGT 120

Db 1861 CTTTTTAACTGACAAATAAGTAAATAATGAGAAAGAAAGAAAGAAAGTGT 1920

Qy 121 CACGTTTGCATATTCGCTATTCAAAATTCGCTTATAGAGTGTATATGTTGGTGGGAC 180

Db 1921 TGTATTTCAATGTTTCGCTATTCAAAATTCGCTTATAGAGTGTATATGTTGGTGGGAC 1980

Qy 181 AATGGCTGCATATCCATGAGGTCGACTCCCTCTCATGATGCTGTTTGTGAAACAT 240

Db 1981 TTTGGCTGCATATCCATGAGGTCGACTCCCTCTCATGATGCTGTTTGTGAAACAT 2040

Qy 241 GACAGATAGCTTTGCAAAATGAGGAAATTTCAAGAAACAAACATTTTCCAGTTTATAATTA 300

Db 2041 GACAGATATCTTTGCAAAATGAGGAAATTTA-----GAAGATCTGATGTCAAACATCAC 2094

Qy 301 TGAAAGTATTAGCAACAATACACAACATTTTCATCAACCATCTGGAGGAGAAATGACAC 360

Db 2095 TAATAGAAAGTATATCAATGATACAGGGTCTTTCATGAATCTGGAGGAGACATGACCAG 2154

Qy 361 GTATGCCCTATTATACAGTGGGATCGGTGTCGGCTGCTGGTGGCTGCTTACATCCAGGT 420

Db 2155 GTATGCCCTATTATACAGTGGGATCGGTGTCGGCTGCTGGTGGCTGCTTACATCCAGGT 2214

Qy 421 TTCATTCTGGTCCCTGGCAGCAGGAGATTCGCTGTTGACGTGATGAGTGGGAGCTTAACAC 480

Db 2215 TTCATTCTGGTCCCTGGCAGCAGGAGATTCGCTGTTGACGTGATGAGTGGGAGCTTAACAC 2274

Qy 481 TGCTATCATGCGACAGGAGATTCGCTGTTGACGTGATGAGTGGGAGCTTAACAC 540

Db 2275 TGCTATCATGCGACAGGAGATTCGCTGTTGACGTGATGAGTGGGAGCTTAACAC 2334

Qy 541 CCGCTCACAGACGATGTCCTCAAAATCAATGAAGAAATTTGGCGACAAAATTTGGAATGTT 600

Db 2335 CCGCTCACAGACGATGTCCTCAAAATCAATGAAGAAATTTGGAATGTT 2394

Qy 601 CTTTCACTCAATAGCAACATTTTTCACCGGTTTATAGTGGGTTTACACGTTGGTGGAA 560

Db 2395 CTTTCACTCAATAGCAACATTTTTCACCGGTTTATAGTGGGTTTACACGTTGGTGGAA 2454

Qy 661 GCTAACCCCTTGTGATTTTGGCCCATCAGCCCTGTTTGGAGCTTTCAGCCGCGCATCTGGGC 720

Db 2455 GCTAACCCCTTGTGATTTTGGCCCATCAGCCCTGTTTGGAGCTTTCAGCCGCGCATCTGGGC 2514

QY	721	AAAGTACTATCTTTCATTTACTGATAAAGAACTCTTGGCCTATGCACAAAGCTGGAGCAGT	780
DB	2515	AAAGTACTATCTTTCATTTACTGATAAAGAACTCTTGGCCTATGCACAAAGCTGGAGCAGT	2574
QY	781	AGCTGAAGAAGCTTTAGCAGCAATCAGAACTGTGATTTGCCCTTTGGAGGACAAAGAAAGA	840
DB	2575	AGCTGAAGAAGCTTTGGCAGCAATTAAGAACTGTGATTTGCCCTTTGGAGGACAAAGAAAGA	2634
QY	841	ACTTTGAAGGTCACAAACAAATTTAGAAGAAAGCTTAAAGGAATTTGGCATATAAGAAAGCTAT	900
DB	2635	ACTTTGAAGGTCACAAACAAATTTAGAAGAAAGCTTAAAGGAATTTGGCATATAAGAAAGCTAT	2694
QY	901	CACGGCCAACTTTCTATTGGTGGCGCTTCTTATGATCTATGCATCATATGCTCTGGC	960
DB	2695	TACAGCCAAATTTCTATAGGTGTGCTTTCTCTGCTGATCTATGCTATCTGCTCTGGC	2754
QY	961	TTTCTGGTATGGGACCTCTTGTGGTCTCTCCAGTGAATATTTCTATTGGCAAGTACTCAC	1020
DB	2755	CTTCTGGTATGGGACCACTTGGTCTCTCTCAGGGGAATTTCTATTGGCAAGTACTCAC	2814
QY	1021	TGTCTTCTTTTCTGTATTAATTTGGGCTTTTGTATTGGACAGGCATTCOCCAAAGCATTTGA	1080
DB	2815	TGTATTTCTTCTGTATTAATTTGGGCTTTTGTATTGGACAGGCATTCOCCAAAGCATTTGA	2874
QY	1081	AGATTTGCAAACCGAAGAGGAGCAGCTTATGAATCTTCAAGTAAATTTGACATAATAAAC	1140
DB	2875	AGCATTTGCAAAATGCAAGAGGAGCAGCTTATGAATCTTCAAGTAAATTTGACATAATAAAC	2934
QY	1141	AAGCATTGACAGCTATTCGAAGAGTGGACATAAAACAGATAATAATTAAAGGAAATTTGGA	1200
DB	2935	AAGTATTGACAGCTATTCGAAGAGTGGGACAAACACAGATAATAATTAAAGGAAATTTGGA	2994
QY	1201	ATTCAAAATGTTCACTTCAGTTACCTTCTCGRAAAGAGTGTAAAGATCTTAAAGGCTCT	1260
DB	2995	ATTCAAAATGTTCACTTCAGTTACCTTCTCGRAAAGAGTGTAAAGATCTTAAAGGCTCT	3054
QY	1261	CAACTCAAGGTTTCAGAGTGGGCAGACAGTGGCGTGTGGGAACAGTGGCTGCGGGAA	1320
DB	3055	GAACCTCAAGGTCAGAGTGGGCAGAGGTGGCCCTGTGTGGAACAGTGGCTGCGGGAA	3114
QY	1321	GAGCAGACCGTGAGCTGATGCAGAGGCTCTATGACCCACAGATGCGCTATGCTGTAT	1380
DB	3115	GAGCAGACAGTCCAGCTGATGCAGAGGCTCTATGACCCACAGAGGGATGTCAGTGT	3174
QY	1381	TGATGGACAGACATTAGGACCAATAATGTAAGGCATCTTCGGGAAATTTACTGTTGGT	1440
DB	3175	TGATGGACAGGATATTAGGACCAATAATGTAAGGTTTCTACGGGAAATCATTTGGTGGT	3234
QY	1441	GAGTCAGAGCGTGTGTTGTTGCCACACAGATAGCTGAAACACATTCGCTATGCCCGGA	1500
DB	3235	GAGTCAGAGCACTGTATTGTTGCCACACAGATAGCTGAAACACATTCGCTATGCCCGTGA	3294
QY	1501	AAATGTCAACATGGATGAGATTTCAGAAAGCTGTAAAGAACCAATGCTCTATGATTTTAT	1560
DB	3295	AAATGTCAACATGGATGAGATTTCAGAAAGCTGTCAAGGAACCAATGCTCTATGATTTTAT	3354
QY	1561	CATGAACACTACCTAATAATTTGACACTCTGGTGTGAGAGAGAGGGGCCAGCTGAGTGG	1620
DB	3355	CATGAACACTGCTCATAAATTTGACACCTTGGTGTGAGAGAGAGGGGCCAGTGTGAGTGG	3414
QY	1621	TGGCAGAAACAGAGAAATCGCCATTTGCTGGGCCCTTGGTTCGCAACCCCAAGATTTCTCT	1680
DB	3415	TGGGCAGAAAGCAGAGGATCGCCATTTGCACTGGCCCTGGTTCGCAACCCCAAGATTTCTCT	3474
QY	1681	GCTGGATGAGGCACGTCAGCTCTGGTGTGAGAGAGAGGGGCCAGCTGAGTGGCCCT	1740
DB	3475	GCTGGATGAGGCCACGTCAGCTCTGGTGTGAGAGAGAGGGGCCAGCTGAGTGGCCCT	3534
QY	1741	GGATGAAGCCAGAAAAGCCGGACTACCAATGTGTGATAGCTCATCGTTTGTCTACAGTTGCG	1800
DB	3535	GGATGAAGCCAGAAAAGGTCGGACCAACATTTGTGTAGTGTCTATCGTTTGTCTACAGTTGCG	3594
QY	1801	TAATGCCGATGTCAATTCGCTGGTGTGTGATGTGAGTCAATTTGTGGAGAAAGGAATTCATGA	1860

3595	Db	TAATGCTCAGCTCATCGCTGGTTTCGATGATGAGTCAATTGGAGAAAGAAATCATGA	3654
1861	QY	TGAACCTCATGAAAGAGAGAGGCATTTACTTTCAAACCTTGTCACAATGCACAGAGAGAAA	1920
3655	Db	TGAACCTCATGAAAGAGAGAGGCATTTACTTTCAAACCTTGTCACAATGCACAGAGAGAAA	3714
1921	QY	TGAAATTCGAGTTAGAAATGCCACTGCTGGTGAATCCAAAAGTGAAGTGATGCCCTTGGAAAT	1980
3715	Db	TGAAATTCGAGTTAGAAATGCCACTGCTGGTGAATCCAAAAGTGAAGTGATGCCCTTGGAAAT	3774
1981	QY	GTCTCAAAAGATTCAAGGTCAGGTCAGGTTTAATAAAAAGAGATCAACTCGCAGGAGTATACA	2040
3775	Db	GTCTCAAAAGATTCAAGTCCAGTCTAATGAAGAAAAGATCAACTCGTAGGAGTGTCCG	3834
2041	QY	TGCACCAAGGCCAAGACAGAAAGCTTGGTACAAAAGAGAGACTTCAATGAGAATGTACC	2100
3835	Db	TGGATCACAAAGCCCAAGACAGAAAGCTTAGTACCAAAGAGGCTCTGGATGAAGATATACC	3894
2101	QY	TGCAGTTTCCTTCTCGAGGATTTCTGAAGCTGAACTCAACTGAATGGCCCTTATTTTGTGT	2160
3895	Db	TCCAGTTTCCTTTCGAGGATTAAGAAGCTAAATTTAACTGAATGGCCCTTATTTTGTGT	3954
2161	QY	TGGTATATTTTGTGCTATTAATAAGGAGGCTCGACACAGACTTTCATATATATTTTC	2220
3955	Db	TGGTGTATTTTGTGCACTATAAATGGAGGCTCGAACACAGACTTTCGAATATATTTTC	4014
2221	QY	AAGGATTAAGGATCTTTACCGAGATGAGGATCCTGAAACAAAACGACAGAAATAGTAA	2280
4015	Db	AAAGATTAAGGGTTTTCACAAAGATTTGATGATCCTGAAACAAAACGACAGAAATAGTAA	4074
2281	QY	CATGTTTTCTGATATGTTTCTAGTCCTTGGAAATTAATTTCTTTTATTAACATTTTCCTCCA	2340
4075	Db	CTTGTGTTTCACTATTTGTTTCTAGSCCTTGGAAATTAATTTCTTTTATTAACATTTTCCTTCA	4134
2341	QY	GGCTTTCAATTTGGCAAAAGCTGGGAGATCCTCACTAAGCGGCTTCGATACATGTTTT	2400
4135	Db	GGGTTTCAATTTGGCAAAAGCTGGAGATCCTCACCAGCGGCTCCGATACATGTTTT	4194
2401	QY	CAGATCCATGCTGAGACAGAGATGTGAGCTGGTTTGATGACCCCTAAAACACCACTGGAGC	2460
4195	Db	CCGATCCATGCTCAGACAGAGATGTGAGTTGGTTTGATGACCCCTAAAACACCACTGGAGC	4254
2461	QY	ATTGCAACCCAGGCTTGCCAATGATCGGCTCAAGTTAAAGGGCTATAGTTCACAGGT	2520
4255	Db	ATTGACTACCGGCTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCACAGGT	4314
2521	QY	TGCTGTCAATTACCCAGAATATAGCAAACTTGGGACAGGCATTTATATACCTTAATCTA	2580
4315	Db	TGCTGTAAATTACCCAGAATATAGCAAACTTGGGACAGGAATAATTATATCCTTCATCTA	4374
2581	QY	TGGTTGCAATTAACACTTTTACTCTTAGCAATTCGTACCCTCATTCGAATAGCAGGAGT	2640
4375	Db	TGGTTGCAACTTAACACTGTTACTCTTAGCAATTCGTACCCTCATTCGAATAGCAGGAGT	4434
2641	QY	TGTTGAAATGAAATGTGTTCTGGACAGCACTGAAAGATGAAAGAGCTAGAGAGNC	2700
4435	Db	TGTTGAAATGAAATGTGTTCTGGACAGCACTGAAAGATGAAAGAGCTAGAGAGTGC	4494
2701	QY	TGGGAAGATTGCTACAGAGCCATCAAAACTTCGAACTGTGTTTCTTTGACTTCGGGA	2760
4495	Db	TGGGAAGATCGCTACTGAAGCAATAGAAACTTCGAACTGTGTTTCTTTGACTTCAGGA	4554
2761	QY	GCAGAAGTTTGAATACATGTTATGCACAGAGTTTGCAGTACCAATACAGAACTCTTTGAG	2820
4555	Db	GCAGAAGTTTGAACATATGTTCTCAGAGTTTGCAGTACCAATACAGAACTCTTTGAG	4614
2821	QY	GAAGACACATCTTCGGGTCTCATTTTCTATCACCCAGGCAATGATGTTTTTCCTA	2880
4615	Db	GAAGACACATCTTTGGAAATTCATTTTCTTCACCCAGGCAATGATGTTTTTCCTA	4674
2881	QY	TGCTGGCTGTTTTCGGGTTTCGGTGCCTACTTGGTGGCAATGATGTTTCAGAACTTCAGGA	2940

```
Db 4675 TGCTGGATGTTTCGGGTTTGGAGCCTACTTGGTGGCACAATAAATCATGAGCTTTGAGGA 4734
QY 2941 TGTCTCTTTGGTATTCACGCTATGCTTTGGTGGCATGCGAGTGGGCGAGTCAGTTC 3000
Db 4735 TGTCTCTTAGTATTTTCAGCTGTGTCTTTGGTGCCATGGCGTGGGCGAAGTCAGTTC 4794
QY 3001 ATTTGCTCTGATGATGCCAAAGCCAAAGATPATCAGCAGCCAGCGTGCATCATGATCATTTGA 3060
Db 4795 ATTTGCTCTGATGATGCCAAAGCCAAAGATPATCAGCAGCCAGCGTGCATCATGATCATTTGA 4854
QY 3061 AAAAGCCCTCTCATTTGACAGCTACAGCCCTCAGCGGCTCAAGCCAAATACGTTGGAAG 3120
Db 4855 AAAAGCCCTCTCATTTGACAGCTACAGCCCTCAGCGGCTCAAGCCAAATACGTTGGAAG 3120
QY 3121 AAATGTCACATTTAATGAGTGTGTTCAACTATCCACTCGACACAGATCCCCGTGCT 3180
Db 4915 AAATGTCACATTTGAGTGTGTTCAACTATCCACTCGACACAGATCCCCGTGCT 3180
QY 3181 CAGGGGCTGAGCCTCGAGGTGAAGGGCCAGAGCCTGGCCCTGCTAGGTAGCAGTGG 3240
Db 4975 TCAGGGACTGAGCCTGAGGTGAAGGGCCAGAGCCTGGCTGCTGTTGGCGAGCAGTGG 5034
QY 3241 CTGTGGGAAGACACAGTTGTTTCAGCTCTAGAGCCCTCTATGACCCCTTGGCTGGTTC 3300
Db 5035 CTGTGGGAAGACACAGTTGTTTCAGCTCTAGAGCCCTCTATGACCCCTTGGCTGGTTC 3300
QY 3301 AGTGTCTAATTTGATGGCAAGAGATAAAGCACCTGAATGTCAGTGGCTCGGAGCACCT 3360
Db 5095 AGTGTCTGCTTGTGGAAGAAATAAAGCGACTGAATGTTAGTGGCTCCGAGCACACT 5154
QY 3361 GGGCATCGTGTCTCAGGAGCCCATCCTGTTTGCATGACGATTCGCGAAGCAATTCGCTA 3420
Db 5155 GGGCATCGTGTCTCAGGAGCCCATCCTGTTTGCATGACGATTCGCGAAGCAATTCGCTA 5214
QY 3421 TGGAGACAACAGCCGGTGTATCATGATCAAGAGATTATGAGGAGCAGCAAGAGGCGCA 3480
Db 5215 TGGAGACAACAGCCGGTGTATCATGATCAAGAGATTATGAGGAGCAGCAAGAGGCGCA 5274
QY 3481 CATACACCACTTCATCGAGACACTCCCTGAGAAATACAAACAGAGTAGGAGACAAAGG 3540
Db 5275 CATACATGCTTTCATCGAGTCACTGCTTAATAATATAGCACTAAGTAGAGAGCAAGG 5334
QY 3541 AACCCAGCTCTCTGTGGCCAGAAACAGCGCATGCCATAGCTCGCTCTTTGTTAGACA 3600
Db 5335 AACTCAGCTCTCTGTGGCCAGAAACAGCGCATGCCATAGCTCGCTCTTTGTTAGACA 5394
QY 3601 GCCTCATATTTTGTCTTTGGATGAAGCTACATCAGCTCTGATACAGAAAGTGAAGGT 3660
Db 5395 GCCTCATATTTTGTCTTTGGATGAAGCCACGTCAGCTCTGATACAGAAAGTGAAGGT 5454
QY 3661 TGTCGAAGAGCCCTGGACAAGCCAGAGAGCGCCACCTGCAATTTGTATGCGCCACCG 3720
Db 5455 TGTCGAAGAGCCCTGGACAAGCCAGAGAGCGCCACCTGCAATTTGTATGCTCACC 5514
QY 3721 CTGTGCCACCATCCAGATGAGATTTAATAGTGTGTTTCAGAAATGGCAAGTCAAGGA 3780
Db 5515 CTGTGCCACCATCCAGATGAGATTTAATAGTGTGTTTCAGAAATGGCAAGTCAAGGA 5574
QY 3781 GCATGGCACACATCAACAGCTGCTGGCCAGAAAGGCAATCTATTTTCCATGGTCAAGT 3840
Db 5575 GCATGGCACAGCATCAGAGCTGCTGGCCAGAAAGGCAATCTATTTTCAATGGTCAAGT 5634
QY 3841 CCAGGCTGGAGCAAGCGCTAGTGAACCTGTGGCCATATGAGCTGTAAATATTTTAAAT 3900
Db 5635 CCAGGCTGGAGCAAGCGCCAGTGAACCTCTGACTGTATGAGATCTTAATACITTTTAAAT 5694
```

RESULT 8

```
US-08-793-610-6
; Sequence 6, Application US/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
```

```
; APPLICANT: STOCKING-HARBERS, Carol
; APPLICANT: OSTERTAG, Wolfram
; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,610
; FILING DATE: 07-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 31 973.8
; FILING DATE: 08-SEP-1994
; PRIOR APPLICATION NUMBER: DE 195 03 952.1
; FILING DATE: 07-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/03175
; FILING DATE: 10-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berman, Richard J.
; REGISTRATION NUMBER: 39,105
; REFERENCE/DOCKET NUMBER: PI614-7007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9318 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
; US-08-793-610-6
```

```
Query Match 75.7%; Score 3239.6; DB 2; Length 9318;
Best Local Similarity 89.6%; Pred. No. 0;
Matches 3495; Conservative 0; Mismatches 399; Indels 6; Gaps 1;
```

```
QY 1 GGAGCGGAGTCCGGATGGATCCTGNAAGAGGCGCGTAAGGGAGTGCAGAGAAGCACTT 60
Db 1760 GGAGCGGAGTCCGGATGGATCCTGNAAGGGGACCGCAATGGAGGAGCAAGAGAGAA 1819
QY 61 CTGAAATAGGCAAAAAAGTAAAAAATGAGAAAGAAAGAAAGAAAGCAACTGTCTAG 120
Db 1820 CTTTTTAAACTGAACAATAAAGTGAAGAAGATAGAGAGAAAGAAAGCAACTGTCTAG 1879
QY 121 CAGCTTTGCAATGTTTCGCTTATTCAAATGGCTTGTAGTGTATATGTTGGTGGGAC 180
Db 1880 TGTATTTCAATGTTTCGCTTATTCAAATGGCTTGTAGTGTATATGTTGGTGGGAAAC 1939
QY 181 AATGGCTGCCATCATCATGAGCTGCACCTCCCTCTCATGATGCTGGTTTGGGAACAT 240
Db 1940 TTTGGCTGCCATCATCATGAGGCTGCACCTCCCTCTCATGATGCTGGTTTGGGAAAT 1999
QY 241 GACAGATAGCTTTGCAATTCAGAGAAATTCAGAAACAAAACTTTTCCAGTTATATAA 300
Db 2000 GACAGATATCTTTGCAATTCAGAGAAATTTA-----GAGATCTCATCTCAACATCAC 2053
QY 301 TGAAGATTTACGAACAATACACAACATTTTCATCAACCAATCTGGAGGAGAAATGACCAC 360
Db 2054 TAATAGAAGTGATATCAATGATACAGGGTTCTTTCATGATCTGGAGGAGAGACATGACCAG 2113
```


Db 4274 TGCTGTAATACCCAGAAATATAGCAAAATCTTGGACAGGAATAATATATCCTTCATCTA 4333
QY 2581 TGGTGGCAATTAACACTTTTACTCTTAGCAATTTGTACCATCATTTCAATAGCAGAGT 2640
Db 4334 TGGTTGGCACTAACACTGTTACTCTTAGCAATTTGTACCATCATTTCAATAGCAGAGT 4393
QY 2641 TGTGAAATGAAATGTTGCTGTGACAAAGCACTGAAAGATAAAGAAAGAGCTAGAAGGAGC 2700
Db 4394 TGTGAAATGAAATGTTGCTGTGACAAAGCACTGAAAGATAAAGAAAGAGCTAGAAGGAGC 4453
QY 2701 TGGGAAATGCTACAGAGCCATCGAAACTTCGGAACCTGTTGTTTCTTTGACTCGGGA 2760
Db 4454 TGGGAAATGCTACTAGGAAATAGAGAACTTCGGAACCTGTTGTTTCTTTGACTCAGGA 4513
QY 2761 GCAGAACTTGAATACATGATGACAGAGTTTGCAGAGTACCATACAGAAACTCTTTGAG 2820
Db 4514 GCAGAACTTGAACATGATGATGCTCAGAGTTTGCAGAGTACCATACAGAAACTCTTTGAG 4573
QY 2821 GAAAGCACATCTTCGGGGTCTCAATTTCTATCACCCAGGCAATGATGTATTTTCCCTA 2880
Db 4574 GAAAGCACATCTTTGGAATTAACATTTCTTCACCCAGGCAATGATGTATTTTCCCTA 4633
QY 2881 TGCTGGCTGTTCCGGTTTGGTGCCTACTTGGTGGCAATGAGTTTCAATGAACTTCAGGA 2940
Db 4634 TGCTGGATGTTCCGGTTTGGAGCCTACTTGGTGGCACATAAACTCATGAGCTTTGAGGA 4693
QY 2941 TGTCTTTTGGTATCTCAGCTATTGTTCTTTGTCATGCGCAGTGGCGAGGTGCTGTC 3000
Db 4694 TGTCTGTTAGTATTTTCACTGTTGTTTGGTGCATGGCGTGGGCAAGTCACTGTC 4753
QY 3001 ATTTGCTCTGATGATGCCAAAGCCAAAGATATCAGCAGCCCGCTCATCATGATCATTTGA 3060
Db 4754 ATTTGCTCTGATGATGCCAAAGCCAAAGATATCAGCAGCCCGCTCATCATGATCATTTGA 4813
QY 3061 AAAAGCCCTCTGATGACAGCTACAGCCCTCAGCGCCCTCAAGCCAAATAGCTTGAAGG 3120
Db 4814 AAAAGCCCTTGTATGACAGCTACAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4873
QY 3121 AAATGTGACATTTAATGAGTCTGTTCAACTATCCACTCGACACAGATCCCGTGTCT 3180
Db 4874 AAATGTGACATTTGTTGAGTTGTTATTTCAACTATCCACCCGACGACATCCAGTGTCT 4933
QY 3181 CAGGGGCTGAGCTCAGGTGAAGAGGCGCAGAGCTGGGCCCTGTTAGGTAGCAGTGG 3240
Db 4934 TCAGGAGCTGAGCTGAGGTGAAGAGGCGCAGAGCTGGCTGTTGGCGCAGCAGTGG 4993
QY 3241 CTGTGGGAAGACACAGTGTTCAGCTCTCAGAGCGCTTATGACCCCTTGGCTGGTTC 3300
Db 4994 CTGTGGGAAGACACAGTGTTCAGCTCTCAGAGCGGTTTCTACGACCCCTTGGCAGGGA 5053
QY 3301 AGTGTAAATGATGGCAAGAGATTAAGCACCTGAATGTCCAGTGGCTCCGAGCACACCT 3360
Db 5054 AGTGTGCTTGTATGGCAAGAAATTAAGAGGACTGAATGTTAGTGGCTCCGAGCACACCT 5113
QY 3361 GGCATCGTGTCTCAGGAGCCCATCGTGTGACTGACGATTCGCGAAGCAATTCGCTA 3420
Db 5114 GGCATCGTGTCCAGAGGCCCATCGTGTGACTGACGATTCGCGAAGCAATTCGCTA 5173
QY 3421 TGGAGAACACAGCGGCTGTATACATGAAGAGATTATGACGAGCAGCAAGAGGCCAA 3480
Db 5174 TGGAGAACACAGCGGCTGTATACATGAAGAGATTATGACGAGCAGCAAGAGGCCAA 5233
QY 3481 CATACACCACTTCATCGAGACACTCCCTGAGAAATACACACAGAGTAGGAGACAAAGG 3540
Db 5234 CATACATGCTTCATCGAGTCACTGCTTAATAATATAGCACTAAAGTAGGAGACAAAGG 5293
QY 3541 AACCCAGCTCTGCTGGCCAGAAACAGGCAATTCGCATAGCTCGCGCTCTTTGATGACA 3600
Db 5294 AACTCAGCTCTGCTGGCCAGAAACAGGCAATTCGCATAGCTCGCGCTCTTTGATGACA 5353
QY 3601 GCCTCATATTTTGGCTTTTGGATGAAGCTTACATCACTCTGGATACAGAAAGTGAAGG 3660
Db 5354 GCCTCATATTTTGGCTTTTGGATGAAGCCACGCTCAGCTCTGGATACAGAAAGTGAAGG 5413

QY 3661 TGTCBAAGAACCCCTGGACAAAGCCAGAGAGGCCACCTGCATGTGATGCGCCACGG 3720
Db 5414 TGTCCAAAGAGCCCTGGACAAAGCCAGAGAGGCCACCTGCATGTGATGCGCCACGG 5473
QY 3721 CTGTGCCACCATCCAGAAATGCAGATTTAATAGTGGTGTTCAGAAATGGCAAGTCAAGA 3780
Db 5474 CTGTGCCACCATCCAGAAATGCAGATTTAATAGTGGTGTTCAGAAATGGCAAGTCAAGA 5533
QY 3781 GCATGSCACACATCAACAGCTGCTGGCCCGAGAAAGGCATCTATTTTCCATGCTCAGTGT 3840
Db 5534 GCATGSCACACATCAACAGCTGCTGGCCCGAGAAAGGCATCTATTTTCCATGCTCAGTGT 5593
QY 3841 CAGGCTGAGCAAAAGCGCTAGTGAACCTGTGGCCATATGAGCTGTAAATATTTTAAAT 3900
Db 5594 CCAGGCTGGAACAAAGCGCCAGTGAACCTGTATGAGATGTAAATATTTTAAAT 5653

RESULT 9

US-09-120-513-1
; Sequence 1, Application US/09120513
; Patent No. 6025160
; GENERAL INFORMATION:
; APPLICANT: Brun, Kimberly
; APPLICANT: Chenery, Richard
; APPLICANT: Ellens, Harma
; APPLICANT: Field, John
; APPLICANT: Yue, Lin
; TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE
; TITLE OF INVENTION: SEQUENCES ENCODING RAT MDR1B2 AND
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY:
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/120,513
; FILING DATE: 22-JUL-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: King, William T
; REGISTRATION NUMBER: 30,954
; REFERENCE/DOCKET NUMBER: GP50008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5015
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4233 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-120-513-1

Query Match 63.2%; Score 2705; DB 3; Length 4233;
Best Local Similarity 81.5%; Pred. No. 0;
Matches 3213; Conservative 0; Mismatches 700; Indels 28; Gaps 6;
QY 5 GCGCAGCTCGGGATGATCTCTGAAGGAGGCCGCTAAGGGGAGTCAGAGAGAACTTCTGG 64
Db 14 CGTGAGGCCCATGGAGTTGAAGAGGGCTTAAACGGAACAGCAGCAAGAACTTCTCA 73

Db 14 CGTAGGCCACCATGGAGTTTGAAGAGGGCCTTAACGGAAACAGCAGACAAGAACTTCTCA 73
Qy 65 AAAATGGCAAAAAGTAAATAAATGAGAAGAAAGAAAGAAACCAACTGTGACGACG 124
Db 74 AAGATGGCAAAAAGAGTAA--AGGAGAGGAGAAACCTGCTGTTGGCATA 127
Qy 125 TTTGCAATTTTCGCTATTCAAAATGGCTTGATAGTGTGTATGTTGGTGAGCAATG 184
Db 128 TTCGGATGTTTCGCTATGCGATGAGTGGCTTGACAGCTGTGCATGGCTCTGGAACTCTC 187
Qy 185 GCTGCATCATCCATGAGGAGCTGCATCCCTCTCATGATGCTGGTGTGGGAAACATGACA 244
Db 188 GCTGCTACTATACACGGCACTGCTCCCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 247
Qy 245 GATAGCTTGCAAAATGCAGGAATTTCAAGAAACAAATTTTCCAGTTATATTAATGAA 304
Db 248 GATAGTTTTACCCAGCAG-----AGACCCGCATCTGCGGAGCGTTACTAATCAA 298
Qy 305 AGTATTACGAACAATACACAACATTTCA---TCAACCATCTGGAGGAGAAATGACCAAG 361
Db 299 AGTGAATCAACASTACACAGACCTCAGCAGACAGCTCTGGAGGAGACATGCCATG 358
Qy 362 TATGCTATTATACAGTGGGATCGGCTGCGCTGCTGGCTGCTGGCTGCTTACATCAGGTT 421
Db 359 TACGCTACTATTACACGGGCAATGGTCCGCTGCTCATGTTGCTCATCATCCAGTT 418
Qy 422 TCATCTGCTGCTGCGCAGCAGAAAGACAGATACTCAAAATTAGAAAACAATTTTTCAT 481
Db 419 TCATTTGGTGGCTGGCAGCTGGGAGACANATACACAGATTAGCCAGAAATTTTCCAT 478
Qy 482 GCTATCATCGCAGAGAGATTTGGCTTGACGTGCGATGAGTGGGAGCTTAAACACC 541
Db 479 GCATCATGAATCAGGAGATAGCTGGTTGACGTGAATGACGCTGGGAGCTCAACACC 538
Qy 542 CGGCTCAGACAGATGTCTCCAAATCAATGAGGAATTTGGCGACAAATTTGGAATGTT 601
Db 539 CGGCTCAGAGATGACGTCTCCAAATTAATGACGGAATTTGGTGACAAATTTGGAATGTT 598
Qy 602 TTTCACTCAATAGCAACATTTTACCGGTTTTATAGTGGGTTTACACGTGGTTGGAAG 661
Db 599 TTTCACTCAATAGCAACATTTTACCGGTTTTATAGTGGGTTTACACGTGGTTGGAAG 658
Qy 662 CTAACCCCTTGATTTGGCCATCAGCCCTGTTCTTTGGACTTTTACGCCCATCTGGGCA 721
Db 659 CTAACCCCTTGATTTGGCCCTCAGCCCTCTTATTTGGGTTGTCATCTGCCATGTGGCA 718
Qy 722 AAGTACTATCTTACTGATTAAGAACTCTTGGCTATGCAAAAGCTGGACGCTA 781
Db 719 AAGGTACTGACTTCAATTAAGAACTCCAGGCTTATGCGAAAGCTGGAGCAGTT 778
Qy 782 GCTGAAGAAGTCTTAGCAGCAATCAGAACTGTGATTGCTTTGGAGGACAAAAGAA 841
Db 779 GCGAAGAAGTCTTAGCAGCAATCAGAACTGTGATTGCTTTGGAGGACAAAAGAA 838
Qy 842 CTTGAAAGTACAAAAAATTTAGAAGAAGCTAAAGGAATTTGGATTAAGAAAGCTATC 901
Db 839 CTTGAAAGTACAAATTAAGAAAGCTAAAGAGTTGGCATAAAGAAAGCCATC 898
Qy 902 ACGGCCAACATTTCTATTTGGTGGCCCTTCTTATTAATGATCATATGCTCTGCTGCT 961
Db 899 ACGGCCAACATTTCCATATGTTATGCTACCTACCTGTTGGTCTATGCTCTTATGCTAGCTGCA 958
Qy 962 TTTCTGATTTGGGACCTCTTGGTCTCTCCAGTGAATATTTCTATTGGCAAGTACTCACT 1021
Db 959 TTTCTGATTTGGGACCTCTTGGTCTCTCAATGAATATTTCTATTGGCAAGTGTACC 1018
Qy 1022 GTCTCTTTTCTGATTAATTTGGGCTTTTGTAGTTATGGACAGGCATCCCCAAGCATGAA 1081
Db 1019 GTCTCTTCTCTATTTATTTGGGACTTTTCAATGATATTTCTATTGGCAAGTGTACC 1078
Qy 1082 GCATTTGCAACCGAAGAGCAGCTTATGAATCTTCAAGATTAATGACATAAACA 1141
Db 1079 GCCTTTTGCAATGCAAGAGGGGCGCCTATGAATCTTCAAGATTAATGATGAGCA 1138

Qy 1142 AGCATTTGACAGCTATTTCGAAGAGTGGACATPAAACCCAGATAATATTAAAGGAAATTTGGAA 1201
Db 1139 AGCATTTGACAGCTTTCACCAAGGAGACACAAACCCAGAGTATAATTTGGAAATTTGGAA 1198
Qy 1202 TTCAAAAATTTTCACATTTACCTTACCTTCTCGAAAGAAAGTTAAGATCTTTAAAGGCTCTC 1261
Db 1199 TTTAAAAATTTTACTTTCAACTACCACTACCAAGTGAAGTAAAGATCTTTGAAGGCGCTC 1258
Qy 1262 AACCTGAAGCTTCAGAGTGGGACAGATGGCGCTGGTTGGGAACAGTGGCTCGCGGAAG 1321
Db 1259 AACCTGAAGTGAAGAGGGGACAGCGTAGCCCTGTTGGCAACAGTGGCTGTGGGAA 1318
Qy 1322 AGCACGACCTGACAGCTGATGACAGAGCTCTATGACCCACAGATGGCATGGTCTGTATT 1381
Db 1319 AGCACAACTCTCCAGCTGCTGCAGAGGCTCTACGACCCCATAGAGGGGAGGTCAATC 1378
Qy 1382 GATGGACAGACATTTAGGACCATTAATGTAAAGCATCTTCGGGAAATTTACTGTGTGGTG 1441
Db 1379 GACGACAGGACATCAGGACCATCAATGTAGGTATCTGCGGAAATCAATTTGGGTGGTG 1438
Qy 1442 AGTCAGAGCTGTGTGTTGTTGGCCACCAGATAGCTGAAACATTCGCTATGCGCCGAA 1501
Db 1439 AGTCAGGAACCGTGTGTTGGCCACCAGATTTGCCAACCATTTGCTATGCGCCGAA 1498
Qy 1502 AATGTCACCATGATGATGAGATTGAGAAAGCTGTTAAGGAAGCCAAATGCTATGATTTTATC 1561
Db 1499 AACGTCACCATGATGATGAGAAAGCTGTCAAGGAAGCCAAATGCTATGACTTTCATC 1558
Qy 1562 ATGAACTACCTAATAATTTGACACTCTGTTGGAGAGAGGGCCGACGCTGAGTGGT 1621
Db 1559 ATGAACTGCCCCACAAATTTAAACACCTGTTGGTGGAGAGGGGCGAGCTGAGTGGG 1618
Qy 1622 GGACAGAAACAGAAATCGCCATTTGCTGGGCGCTGTTGCGCAACCCCAAGATTTCTCTG 1681
Db 1619 GGACAGAAACAGAGATCGCCATTTGCCGGCGCTGTTCCGCAACCCCAAGATTTCTCTG 1678
Qy 1682 CTGGATGAGCAACGCTCAGCTCTGGACCTGGAACCTGAAAGTGAAGCAGTGGTTCAAGTGGCCCTG 1741
Db 1679 TTGGATGAGGCCAGCTCAGCTTTGGACACAGAAAGCCGCTGTTCAAGGCGCTCTG 1738
Qy 1742 GATAAGGCCAGAAAGCCGAGCTACCATTTGATAGCTCATCTGTTGCTACAGTTCGT 1801
Db 1739 GATAAGCTAGAGAGGCCGAGCCACCATTTGTGATAGCTCACCCTGTTGCTACAGTGGC 1798
Qy 1802 AATCCGATGCTATGCTGTTGATGATGAGATCAATTTGGAGAAAGGAATCATGAT 1861
Db 1799 AATGCTGACGCTATGCTGTTGATGTTGTTGCTCATTGAGAGCAAGGAATCATGAA 1858
Qy 1862 GAATCATGAAGAAGAGGCGATTTACTTCAAACCTTGTCAACATGACACAGACTAGAGGAAAT 1921
Db 1859 GAGCTCATGAAGAAGAGGCGATTTACTTCAAACCTTGTCAACATGACACAGACTAGAGGAAAT 1918
Qy 1922 GAAATTTAGTTAGAAATGCCACTGGTGAATCCAAAAGTGAAGTGAATGCTTTGGAATG 1981
Db 1919 GAAATTTAGCAAGAAATAATGCTTATGAATCCCAAGTGAAGTGAATGCTTTGAGTGG 1978
Qy 1982 TCTCCAAAAGATTCAGGCTCCAGTTTAAATAAAGAAGATCAACTCGCAGAGTATACAT 2041
Db 1979 ACTTCAGAAATCAAAATCTCCTTAAT---AAGGAGATCAATTCGACAGATATCCAC 2035
Qy 2042 GCACCAAGGCCAAGACAGAAAGCTTGGTACAAAAGAGCACTTGAATGAGATGTAACCT 2101
Db 2036 AGAAGACAAGACCAGGAGAGAAGACTTGTTCGAAAGAGGATGTTGGAATGAGATGTGCT 2095
Qy 2102 CCAGTTCTCTCGAGGATTTCTGAAGCTCAACTGAATGGCCCTTATTTTGTGTT 2161
Db 2096 ATGGTTCTCTTTGGCAGATCCTAAGACTAATAATTTAGTAAGTGGCCCTATTTAGTTGTG 2155
Qy 2162 GGTATATTTTGTGCTATTATAACGGGCGCTGCAACAGCAATTTTCAATAAATTTTCA 2221
Db 2156 GGTGTACTTTGTGCTTTATAAATGGGTGCATACAACAGTGTGTGGCATAGTGTTTTCA 2215

*_Qy	2222	ACGATTATAGGATCTTTACCCGAGATGAGGATCTTGAAACAAACAGACAGAAATAGTAAAC	2281
Db	2216	AAGATTGTAGGGGTTTTTCAAGAGACGACGACCATGAAACCAACAAACGGAATGTTAAAC	2275
Qy	2282	ATGTTTTCTGTATTTGTTTCTAGTCTTGGAATTAATTTCTTTTATATACATTTTCTCCTCAG	2341
Db	2276	TTGTTTTCCCTTCTCTTCTGGTCATGGGAATGATTTCTTTTGTTCAGTACTTCTTTTCAA	2335
Qy	2342	GGCTTCACATTTGGCAAAAGCTGGGGAGATCCTCACTAAGCGCTTCGATACATGGTTTTC	2401
Db	2336	GGCTTCACATTTGGCAAAAGCTGGGAGATCCTCACCACAGGACTCCGATACATGTCCTTC	2395
Qy	2402	AGATCCATGCTGACACAGGATGTCAGCTGGTTGTATGACCTTAABAAACACCACTGGAGCA	2461
Db	2396	AAATCCATGCTGACACAGGATTAAGCTGGTTGTATGACCATTAABAAACACCACTGGTCG	2455
Qy	2462	TTGACACACGAGGCTTGCCAAATGATCGGGCTCAAGTTTAAAGGGCTATAGGTTCCAGGCTT	2521
Db	2456	CTGACTACCAAGGCTTCGCTAGTGCAGCTTCTAATGTTTAAAGGGCTATGGCTCCAGGCTT	2515
Qy	2522	GCTGTCAATACCCAGATATAGCAATCTTTAGCAATTTGACCAATGTACCATCATTTGCAATAGCAGGAGTT	2581
Db	2516	GCTGTAGTTACCCAGAATGTAGCAAACTTTGGCACAGGAATTTATCTTATCTTATAGTCTAT	2575
Qy	2582	GGTTGGCAATTAACACTTTTACTCTTTAGCAATTTGACCAATGTACCATCATTTGCAATAGCAGGAGTT	2641
Db	2576	GGCTGGCAGCTTACACTTTTACTTTAGTAATTAATACCACCTCATTTGCTTGGTGGGAATT	2635
Qy	2642	GTTCAAAATGAATGTTGCTGCACAAGCACTGAAAGATAGAAGAGCTAGAAGAGCT	2701
Db	2636	ATTGAAATGAACACTTTGCTGCCTCAAGCTTTGAAGCAACAAGAAAGCTAGAGATCTCT	2695
Qy	2702	GGGAAGATTGCTACAGAAGCCATCGAAAACCTTCCGAACCTTGTTGTTCTTTCACCTCGGAG	2761
Db	2696	GGGAAGATCGCTACAGAAGCAATTGAAAACCTTCCGCACTGTTGCTCTTTGCTGCTAGGAG	2755
Qy	2762	CAGAAGTTTGAATACATGATGACAGATTTGCAAGTACCATACAGAAACTCTTTGAGG	2821
Db	2756	CAGAAGTTTGAATACTATGATGCCAGAGCTTGCAGATACCATACAGAAATGCTTTGAAG	2815
Qy	2822	AAAGCACACATCTTCGGGGTCTCATTTTCTATCACCCAGGCAATGATGATTTTTTCCCTAT	2881
Db	2816	AAAGCACACGCTTTGGGATCACCCTCGCCTTCACCCAGGCCATGATTTATTTTTCTCTAT	2875
Qy	2882	GCTGGCTGTTTTCCGGTTTGGTGCTACTTTGGTGGCAATAGTTCATGAACCTTTCAGGAT	2941
Db	2876	GCTGCTTTGTTCCGGTTTCGGTGCTTACTTGGTGCCATGGCAGTGGGGCAGGTCAGTTCA	2935
Qy	2942	GTTCTTTTGGTATTTCTCAGCTATTGCTTTGGTGCCATGGCAGTGGGGCAGGTCAGTTCA	3001
Db	2936	GTTATGTTGGTATTTTCTGCTGTGTGCTTTGGTGCCATGGCAGGAGGGAATACACGTTCA	2995
Qy	3002	TTTGCTCTCTGACTATGCCAAAGCCAAAGTATCAGCAGGCCACGCTCATCATGATCATTTGAA	3061
Db	2996	TTGCTCTCTGACTATGCCAAAGCCAAAGTCTCAGCATCCACATCATCAGGATCATTTGAG	3055
Qy	3062	AAAAGCCCTCTGATTCAGACGCTACAGCCCTCAGGCCCTCAAGCCAAATACTTTGGGAAGGA	3121
Db	3056	AAAATCCCGAGATTGACAGCTACAGAGGAGGAGGCTTGAAGCCCTAATTTGTTTGAAGGA	3115
Qy	3122	AATGTGACATTTAATGAGTCTGTTTCAACTATCCACTCGACACAGACATCCCGCTGCTC	3181
Db	3116	AATGTGAAATTTAATGAGTCTGTTTCAACTATCCACCCGACCCAAACATCCCAAGTCTT	3175
Qy	3182	CAGGGCTGAGCCTCGAGGTGAAGAGGCCAGACGCTGGCCCTTCGTAGGTAGCAGTGGC	3241
Db	3176	CAGGGACTGAGCTTCGAGGTGAAGAGGGGCAACGCTTGCCTTGGTGGGCACGACGTGGC	3235
Qy	3242	TGTTGGGAAGACACAGTTTCTCAGCTCCTTACAGCGCTTCTATGACCCCTTGGCTGGTTTCA	3301
Db	3236	TGCGGGAGACGATACAGTGTCCAGTGTCCAGGCTGCTTCAGAGCGCTTCTACAACCCCATTTGGCTTGAACAA	3295
Qy	3302	GTCTCAATTTGATGGCAAGAGATAAAGCAACCTGAATGTCAGTGGCTTCCGAGCACACCTTG	3361

[illegible]

```

RESULT 11
US-08-461-823-1
; Sequence 1, Application US/08461823
; Patent No. 5593840
; GENERAL INFORMATION:
; APPLICANT: Bhatnagar, Satish K.
; APPLICANT: George Jr., Albert L.
; APPLICANT: Nazarenko, Irina
; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OncorPharm, Inc.
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,823
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/168,621
; FILING DATE: 16-DEC-1993

```

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,433
; FILING DATE: 27-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Karta, Glenn E.
; REGISTRATION NUMBER: 30,649
; REFERENCE/DOCKET NUMBER: PA-0012 CIP 2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301 527-2058
; TELEFAX: 301 208-6997
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-461-823-1

Query Match 49.5%; Score 2118; DB 1; Length 2726;
Best Local Similarity 87.5%; Pred. No. 0;
Matches 2409; Conservative 0; Mismatches 305; Indels 38; Gaps 7;

QY 1519 GATTGAGAAAGCTGTTAAGGAAGCCCAATGCCTTATGATTTTATCATGAAACTACCTAATAA 1578
Db 1 GATTGAGAAGCTGTCAGGAAGCCCAATGCCTTATGATTTTATCATGAAACTACCTAATAA 60

QY 1579 ATTTGACACTCTGGTTGGAGAGAGGGGCCCGACGTGAGTGGTGACAGAAACAGAGAAT 1638
Db 61 ATTTGACACCCCTGGTTGGAGAGAGAGGGGCCCGACGTGAGTGGTGACAGAAACAGAGAAT 120

QY 1639 CGCCATTGCTCGGGCCCTGGTTGCGAACCCTCAAGATCTTCTGCTGGATGAGGCAACGTC 1698
Db 121 CGCCATTGACCTGGCCCTGGTTGCGAACCCTCAAGATCTTCTGCTGGATGAGGCCACGTC 180

QY 1699 AGCTCTGGACACTGAAAGTGAAAGCAGTGGTTTCAGGTGGCCCTGGATAGAGCCAGAAAAGG 1758
Db 181 AGCCTTGGACACAGAAAGCAGTGGTTTCAGGTGGCCCTGGATAGAGCCAGAAAAGG 240

QY 1759 CCGGACTACATTGATAGCTCAGTCATGCTTGTCTACAGTTCTGAATGCCGATGCTATGCC 1818
Db 241 TCGGACCAACCATTTGATAGCTCAGTCATGCTTGTCTACAGTTCTGAATGCCGATGCTATGCC 300

QY 1819 TGGTTTGTATGATGGAGTCAATTTGTGAGAAAGGAATCATGATGAACATCATGAAAGAGAA 1878
Db 301 TGGTTTGTATGATGGAGTCAATTTGTGAGAAAGGAATCATGATGAACATCATGAAAGAGAA 360

QY 1879 GGGCATTTTACTTCAAACTTTGTCAAAATGCGACAGAGAGAAATGAAATGAGTTAGAAA 1938
Db 361 AGGCATTTTACTTCAAACTTTGTCAAAATGCGACAGAGAGAAATGAAATGAGTTAGAAA 420

QY 1939 TGGCACTGGTGAATCCAAAAGTGAAGTATGCTTTGGAATGCTCTCCAAAAGATTCAGG 1998
Db 421 TGGCACTGGTGAATCCAAAAGTGAAGTATGCTTTGGAATGCTCTCTCAAAATGATTCAG 480

QY 1999 GTCCAGTTTAAATAAAGAGATCAACTCGCAGGAGTATACATGCACCAAGAGCCCAAGA 2058
Db 481 ATCCAGTCTTAATAAAGAGATCAACTCGTAGGAGTGTCCGTGGATCATCAAGCCCAAGA 540

QY 2059 CAGAAAGCTTGGTACAAAAGAGACATTGAATGAGAAATGTACCTCCAGTTTCTCTCGGAG 2118
Db 541 CAGAAAGCTTGGTACAAAAGAGACCTTGGATGAAAGTATACCTCCAGTTTCTCTTTGGAG 600

QY 2119 GATTCTGAAGCTGAAGTCAACTGAAGTGGCTTATTTGTTGGTGGTATATTTTGTGCTAT 2178
Db 601 GATTATGAAGCTAAATTAAGTGGGCTTATTTGTTGGTGGTATTTTGTGCTAT 560

QY 2179 TATAACCGGAGGCTGCAACACAGCATTTTCAATAATATTTTCAAGGATTATAGGGATCTT 2238
Db 561 TATAATGGAGGCTGCAACACAGCATTTTCAATAATATTTTCAAGGATTATAGGGGTTTT 720

QY 2239 TACCCGAGATGAGGATCCTGGAACAAAACAGACAGAAATAGTAACATGTTTCTGTATTGTT 2298
Db 721 TACAAGAAATTGATGATCCTTGAAACAAAACAGACAGAAATAGTAACATGTTTCTGTATTGTT 780

QY 2299 TCTAGTCCCTGGAAATTTATTTCTTTTATACATTTTCTCCAGGGCTTCCACATTTGGCAA 2358
Db 781 TCTAGCCCTTGGAAATTTATTTCTTTTATACATTTTCTCCAGGGCTTCCACATTTGGCAA 840

QY 2359 AGCTGGGGAGATCCTCACTAAGCGCTTCGATACATGTTTTCAGATCCATGCTGAGACA 2418
Db 841 AGCTGGAGAGATCCTCACTAAGCGCTTCGATACATGTTTTCGATCCATGCTGAGACA 900

QY 2419 GGATGTCAGCTGGTTGATGACCCCTAAAACACACCTGGAGCAATTCACAAACAGGCTTGC 2478
Db 901 GGATGTAGTGGTTGATGACCCCTAAAACACACCTGGAGCAATTCACAAACAGGCTTGC 960

QY 2479 CAATGATGGCTCAAGTTAAAGGGCTATAGGTTTCCAGGCTTGTCTCATTTACCCAGAA 2538
Db 961 CAATGATGGCTCAAGTTAAAGGGCTATAGGTTTCCAGGCTTGTCTCATTTACCCAGAA 1020

QY 2539 TATAGCAAACTCTGGGACAGGCATTTATATATCTTAACTATGTTGGCAATTAACACT 2598
Db 1021 TATAGCAAACTCTGGGACAGGCATTTATATATCTTAACTATGTTGGCAATTAACACT 1080

QY 2599 TTTACTCTTAGCAATTTGATCCCATCATTCGAATGACAGGAGTTGTTGAATGAAATGTT 2658
Db 1081 GTTACTCTTAGCAATTTGATCCCATCATTCGAATGACAGGAGTTGTTGAATGAAATGTT 1140

QY 2659 GTCTGGCAACGACTGAAAGATAGAAAGAGCTAGAAAGAGCTGGGAAGATGCTCTACAGA 2718
Db 1141 GTCTGGCAACGACTGAAAGATAGAAAGAGCTAGAAAGAGCTGGGAAGATGCTCTACAGA 1200

QY 2719 AGCCATCGAAACCTTCGAACTGTTGTTTCTTTGACTCGGGAGCAGAAAGTTTGAATACAT 2778
Db 1201 AGCAATGAAACCTTCGAAACCTGTTTCTTTGACTCAGGAGCAGAAAGTTTGAACATAT 1260

QY 2779 GTATGACAGAGTTTGAAGTACCATACAGAACTCTTTGAGGAAGACACATCTCTCGG 2838
Db 1261 GTATGCTCAGAGTTTGAAGTACCATACAGAACTCTTTGAGGAAGACACATCTCTCGG 1320

QY 2839 GGTCTCATTTTCTATCACCCAGGCAATGATGTAATTTTCTCTGCTGGCTGTTTCCGGTT 2898
Db 1321 AATTACATTTTCTTCCCTCCAGGCAATGATGTAATTTTCTCTGCTGGATGTTTCCGGTT 1380

QY 2899 TGGTGCCTTACTTGGTGGCAATGAGTTTCAGCAACTTTTCAGGATGTTCTTTTGGTATTCTC 2958
Db 1381 TGGAGCCTTACTTGGTGGCAATGAGTTTCAGCAACTTTTCAGGATGTTCTTTTGGTATTCTC 1440

QY 2959 AGCTATTGCTTTGCTGCCATGCGCAGTGGGCGAGTTCAGTTCAATTTGCTCCTGACTATGC 3018
Db 1441 AGCTGTTGCTTTGCTGCCATGCGCAGTGGGCGAGTTCAGTTCAATTTGCTCCTGACTATGC 1500

QY 3019 CAAAGCCAAAGTATCAGCAGCCCGCAGTCAATGATCATTTGAAAAAGCCCTCTGATTGA 3078
Db 1501 CAAAGCCAAAGTATCAGCAGCCCGCAGTCAATGATCATTTGAAAAAGCCCTCTGATTGA 1560

QY 3079 CAGCTACAGCCCTCAGCGCCCTCAAGCCAAATACCTTGAAGGAAATGTGACATTTAATGA 3138
Db 1561 CAGCTACAGCCCTCAGCGCCCTCAAGCCAAATACCTTGAAGGAAATGTGACATTTAATGA 1620

QY 3139 GGTCTGTTTCAACTATCCCACTCGACACAGACATCCCGTGTCTCCAGGGGCTGAGCCCTGA 3198
Db 1621 AGTTGATTTCAACTATCCCACTCGACACAGACATCCCGTGTCTCCAGGGGCTGAGCCCTGA 1680

QY 3199 GGTGAAGAGGGCCAGAGCGTGGCCCTCGTAGGATAGCAGTGGCTGTGGGAAGAGACACAGT 3258
Db 1681 GGTGAAGAGGGCCAGAGCGTGGCTCTGTGGGAGCAGTGGCTGTGGGAAGAGACACAGT 1740

QY 3259 TGTTGAGCTCTAGAGCGCTTCTATGACCCCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3318
Db 1741 GGTCCAGCTCTGAGAGCGGTTCTAGACCCCTTGGCAGGAAAGTGTGCTGCTGCTGCTGCTGCTGCT 1800

QY 3319 AGAGATAAAGCAGCTGAATGTCTAGTGGCTCCGAGCAGACACCTGGGCAATGCTGCTCAGGA 3378

QY 831 AAAAGAAAGAACTTGAAAGGTACAACAAAAATTTAGAGAGAGCTTAAAGGAATTTGGGATAA 890
DB 944 AGGATAAGCTCGCCNAGCAATACGNAACCCATCTGGCTGAGGCTGAAAAATGGGGCGTCA 1003
QY 891 AGAAAGCTATCACGGGCCAACATTTCTATTGTGGCGCTTTCTTATTGATCATCATCAT 950
DB 1004 AACAAAGGCTATCCTTTGGTATGATGATTTGGTGTATGTTCTCGGTATCATGTTCTCGAAT 1063
QY 951 ATGCTCTGGCTTTCTGGTATGGGACCTCCTTTGGTCTCTCCAGTGAATATTCTATTGGAC 1010
DB 1064 ATGGCTCGGTTTGGATGGGATCTCGCTTCGTTGTCGGTAAAGAGTCAACGTGGGCC 1123
QY 1011 AAGTACTCACGTCTCTCTTTCTGTTAATTTGGGGCTTTTAGTATTGGACAGCATCCC 1070
DB 1124 AAGTTCTGACAGTTTGTATGTTCTATCTCTGATCGGTTCTGTTGAGTTGGGCAACGTGGGCC 1183
QY 1071 CAAGCATTTGAAGCATTTGCAACGCAAGAGGAGGAGCTTTTGAATCTTCAAGATATTTG 1130
DB 1184 CCAATGGTCAGGCCCTTTACGAATGGTGTGCTGGCGCCGCAAGATTTTACAGCACGATTTG 1243
QY 1131 ACAATAAACCAAGCATTTGACAGCTATTTCGAAGAGTGGACATAAAACAGATATAATTAAGG 1190
DB 1244 ACCGAGATCGCCACTGGACCCCTATTCTGACGAAGGAAGGTACTCGACCACTTTTGNAG 1303
QY 1191 GAAATTTGGAATTTCAAAATGTTCACTTTCAGTTACCTTCTCGAAAAGAAAGTTAAGATCT 1250
DB 1304 GAAATTCGAATTTTCGCAATGTCAAACACATCTACCTTTCGAAGCCCGCAAGTTACAGTCA 1363
QY 1251 TAAAGGTTCTCAACCTGAAGTTTACAGTTGGGACAGAGTGGCGCTGGTTGGGACAGTG 1310
DB 1364 TGGAAAGATGTCCTTTATCGATGCCCCGCCGGAAGACTTACGCCATTTGGTGGGCCCATCTG 1423
QY 1311 GCTGGGGAAGAGCAGCAGCCGTCAGCTGATGCAGAGGCTCTATGACCCACAGATGGCA 1370
DB 1424 GCTCTGGAAGAGTACTGTTGTCGGCTGGTGGAGCGCTTTTACCTTCCAGTAGGAGGCC 1483
QY 1371 TGGTCTGTATTGATGACAGGACATTAGGACCATTAATGTAGGCACTCTTCGGGAAATTA 1430
DB 1484 AGGTATTGCTGGACGCCATGATATCCAAACCCCTCAACCTCCGTTGGCTGGGACAGAGA 1543
QY 1431 CTGGTGTGTGATCAGGAGCTGTGTTGTTGCCACAGCATAGCTGAATAACATTCGCT 1490
DB 1544 TCTCTCTTGTACGCCAGGAACCTGTCTTTTTCAGCACCACGATCTTTTGAACATTCGAAC 1603
QY 1491 ATGGCCG-----CGAAAATGTCACCATGATGAGA-----TTG 1523
DB 1604 ATGGCTTGATTGSCACCAAAATTCGAGCATAGTCGAAGSACAAGATTAGAGAGCTCGTTG 1663
QY 1524 AGAAGCTGTTAAGGAAGCCATGCGCTATGATTTTATCATGAACACTACCTAATAAATTG 1583
DB 1664 AGAATCGGCCAGAAATGGCCAAATGCTCATGATTTTATTATGCTCTGCTGAAAGTTACG 1723
QY 1584 ACACCTCTGGTGGAGAGAGAGGGGCCACGCTCAGTGGTGGACAGAAACAGAGAATCGCCA 1643
DB 1724 ATACGAATGTGGGTACGCTGTTTCTTACTTTAGAGGTCAGAGGTCAGAAACACGATTTGCCA 1783
QY 1644 TTGCTCGGGCCCTGGTTTCGAACCCCAAGATTTCTTCTGCTGATGAGGCAACGTCAGCTC 1703
DB 1784 TTGCTCGTCCCATTTGTCAGTGACCCCAAGATTTCTGTTGCTTGATGAAGCTACATCAGCTT 1843
QY 1704 TGGACACTGAAGAGTGAAGCAGTGGTTTACGTTGGCCCTGATGAAGGCCAGAAAGSCCGA 1763
DB 1844 TGGATACCAAGTCCGAGGGCGTCTCAACGCGCTCTTGATAAAGCTGCCGAGGAGTAGAA 1903
QY 1764 CTACCAATGTGATGCTCATGCTTCTGCTACAGTTGCTGAATGCCGATGTCATTGCTGGTT 1823
DB 1904 CTACCAATGTGATGCTCATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1963
QY 1824 TTGATGATGAGGATCATTTGGGAAAGAAATCATGATGAATCATGATGAAGAAAGGCA 1883
DB 1964 TGGTGGCGGGAAGATTTGCCGAACAGGGAACACATGACGAATTTGGTCTGATCGCAAGGCA 2023
QY 1884 TTTTACTCAAACTGTCACATGTCACAAAGAGGAATGAATTTAGATTAGAAATGCCA 1943

DB 2024 CGTACTATAAACTTGTGGAGGCGCAACGCTATCAACGAGGAGAGGAAGCAGAACTCTCGG 2083
QY 1944 CTGGTGAATCCAAAAGTGAAAGTGATGCTTGG-----AAATGCTCTCCAAAAGATTTCAGG 1998
DB 2084 AAGCCAGCCGACATGAGGCGCGATGATTTGGTCAAGAAGGGTTACTCGCATCAAGA 2143
QY 1999 GTCCAGTTTAAATAAAAGAAGATCAACTCGCAGGAGTATACATGACCAACAGGCCCAAGA 2058
DB 2144 CTGCAGTTAGCAGCTCGAATTTCTCGATGCTGTAGATGAGAAGGCGCTTTGGAGATGA 2203
QY 2059 CAGAAAGCTTGGTACAAA-----GAGG 2081
DB 2204 AGCGAAACCGGAACGCAAAAATCGGTTTCGAGTGCCTTCTATCTAAGAAGTCCCGCAGC 2263
QY 2082 ACTTGAATGAGATGTACCTCCAGTTTCTCTGAGGAGTTCTGAAAGCTGAACCTCAACTG 2141
DB 2264 AGTTCGAGAGTATTCACTCTGGACCTCGTCAAGTTTCATCGGTGCATTTAACCGCCCTG 2323
QY 2142 AATGGCCCTTATTTTGGTGGTATATTTTGTGCTATATTAACAGGAGGCGCTGCAACCAG 2201
DB 2324 AGCTTGGTTACATGCTCAATTGGCTTGACTTCTCTGCTTCTGCTTGGTGGTGAATTCG 2383
QY 2202 CATTTCAATAATATTTCAAGGATTATAGGA-----TCTTTACCCGAGATGAGATCC 2256
DB 2384 CCAGAGCTTTTCTGTAGCCCAAGGCAATTAGCAGCTCTCGTTACCCGATCAATGTTT 2443
QY 2257 TGAACAAAAACGACAGAAATAGTAACATGTTTCTGTATTTTCTGTATTTCTAGTCTTGGAAATTA 2316
DB 2444 ACAAACTCA--GGCATGACGCGAATTTCTGGTCTTGTATGTTCTCTGTTGGTGAATTCG 2501
QY 2317 TCTTTTATTAACATTTTCTCCAGGCTTCACATTTGGCAAGCTGGGAGATCTCTAC 2376
DB 2502 TCAATTTATCAGCCTGTCTATCAATGGTACAGCATTTGCTATTGTTGGGAGAGACTCAT 2561
QY 2377 TAAAGCGCTTCGATACATGTTTTCAGATCCATGCTGAGACAGGATGTCAGCTGGTTGA 2436
DB 2562 TCGCCGAGCTTAGAAGTCAAGGTTTAGATCGATTCTTCGTCAGSACATCTCAATTTTCGA 2621
QY 2437 TGACCCCTTAAACACCACTCGGAGCTGAGCAACGAGCTTGGCAATGATGCGGCTCAAGT 2496
DB 2622 CAGGGAAGAGAACAGCAGCAGCGTCTGAGCTGCTTCTTATCAACAGAGACGAAGAATCT 2681
QY 2497 TAAAGGGCTATAGTTCCAGGCTGCTGCTTACATACCCAAATATAGCAATCTTGGGAC 2556
DB 2682 GTCCGGGCTTAGCGGAGTGACTCTCGGCACATCATCATGCTAGCACTACGTTGGGGC 2741
QY 2557 AGGCATTTATATATCTTAAATCTATGTTGGCAATTAACACTTTTACTCTTAGCAATGT 2616
DB 2742 GGCGATGATCATTCGATTGGCGATCGGATGGAAGCTGGCTCTGTTTGCATTTCTGTCTGT 2801
QY 2617 ACCCATCATTCGAATAGCAGAGTGTGTTGAATGAATAATGTTGCTGCGACAGCAGTGA 2676
DB 2802 TCCAATCTCTGCGATGCGGCTTCTCAGATTCTACATGCTTGTCTCAATTTCCAGCAAG 2861
QY 2677 AGATAAGAAAGAGCTAGAGGAGCTGGAGATGCTACAGAAAGCCATCGAAAATCTCCG 2736
DB 2862 ATCGAAGTCTGCCCTACGAAGGGTCTCGGAGCTATGCTGGGCGTTTTACCACAGCAGTACA 2921
QY 2737 AACTGTTGTTTCTTGTAGCTCGGAGCAGAGTTTGAATACATGATATGACAGAGTTTGA 2796
DB 2922 CACTGTAGCATCACTCACTCGGAACAAGATGCTGGGCGTTTTACCACAGCAGTACA 2981
QY 2797 AGTACCATACAGAAACCTCTTTGAGGAAAGCACATCTTCGGGGTCTCAATTTCTATCAC 2856
DB 2982 AAAACAGGAGCGGAAGAGTTTGATCTCAGTGTGAGATCTCCCTGCTGTATGCTCGTC 3041
QY 2857 CCAGCAATGATGATTTTCTCTATGCTGGCTGTTTCCGGTTCGCTTACTTGTGGG 2916
DB 3042 GCAGCATTTGGTGTGTTTCTCGCTCGCTTTGGGCTTCTGGTATGCTGTACACTTCTAGG 3101
QY 2917 AAATGAGTTTCACTTTCAGGATGTTCTTTTGGTATTTCTCAGCTATTTGCTTGTGTC 2976


```

; Db 3102 CCATCATGAGTACAGCATCTCCGCTTCTCGTCTGCTTTCTGAGATCTTTTGGTGC 3161
;
; QY 2977 CATGGCAGTGGGAGAGTCACTTTGCTCTGATATGCCAAAGCCAAAGATACAGC 3036
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3162 GCAATACAGCGGAAGTCTCTCTCTTTTGGCCGAGATGGGTAAAGCAAGATGCCG 3221
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3037 AGCCACAGCTCATCATGATCATTTGAAAAAGCCCTCTCATTTGACAGCTACAGCCCTACGG 3096
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3222 TGCTCAATTCAGAAACTCTTCGACAGCAAGCCACCATTTGACATCTGGTCGATAGGG 3281
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3097 CCTCAAGCAAAATACGTTGAAGGAAATGTGACATTTAATGAGTTCGTGTTCAACTATCC 3156
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3282 CGAGAAGTTGGAGTCTATGGAAGCGAAATCGAATTCGGGAGCTCCACITTTAGGTACCC 3341
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3157 CACTCGACACAGATCCCGTGTCTCCAGGGGCTGAGGCTCGAGGTGAAGAAGGCCAGAC 3216
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3342 AACCGCGCGAGAGAGCTGTCTTCGAGGACTGAATTTGAGCGTGAAGCGCTGGACAATA 3401
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3217 GCTGGCCCTCTAGGTAGCAGTGGCTGTGGGAAGAGCACAGTTGTTTCAGCTCCTAGAGCG 3276
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3402 CATTTGCCCTTTGGACCCAGTGGATCGGTAGAGACATACGATTCGCTCTGCTTGAGCG 3461
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3277 CTTCATGACCCCTTGGCTGGTTCAGTGTCTAATTTGATGGCAAGAGATAAGACACCTGAA 3336
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3462 ATTTTATGACGCACTTCTGGAGGGTCTTCGTTGACGGAAGGACATTTACCAACTCAA 3521
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3337 TGTCCAGTGGCTCGAGACACACCTTGGGCATCTGTCTCAGGAGCCCATCTCTGTTGACTG 3396
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3522 TGTCAACTCATACCGCACTTCTCTCTCCCTGTGACGCAAGAACCTACTCTGTATCAGGG 3581
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3397 CAGCATTTGCCGAGAACATTCGCTATGAGGACACACAGCCGGTCTGTATCATCAATGAAGAT 3456
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3582 TACCATCAAGGAAATATCTGCTTGGAGTCGATAGGATGACGTTTCGGAGGAGACTTT 3641
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3457 TATCAGGACGCCAAGAGGCCAACATACACACTTCATCGAGACACTCCCTGAGAAATA 3516
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3642 GATTAAGTCTGCAAGATGCCAATCTATGATTTGTTATGTCACCTCCCTGAGGGATT 3701
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3517 CAACACAGAGTAGGACACAAAGGACCCAGCTCTCTGTTGGCCAGAAACAGGCAATGC 3576
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3702 TGACACCGTGTGGCAGCAAGGAGGCATGTTGCTGTGGTGGACAAACAGCGTCTGC 3761
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3577 CATAGCTCGGCTCTCTGTAGACAGCCTCATATTTTCTTTTGGATGAAGCTACATCAGC 3636
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3762 CATTTGCTGTGCTCTCTGCTGACCCCAAGTCCCTTCTCTGTGATGAAGCCACATCTGC 3821
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3637 TCTGGATACAGAAAGTGAAGGTTGTCCAAGAGCCCTCGACAAAGCCAGAGAGGCCG 3696
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3822 TCTTGACTCCGAATCTGAGAAAGTCGTACAAGCTGCGCTGGATGCTGCTGCCCGCGGCG 3881
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3697 CACCTGATTTGTATCCGCCACCGCTTGTCCACATCCCAANTGCAGATTTAATAGTGT 3756
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3882 GACAAAGATTGCTGTTGCCCAAGGCTGAGCACCATTTCAAAAGCTGATATAATTTACGT 3941
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3757 GTTTCAGAAATGGCAAGTCAAGAGCATGGCACACATCAACAGCTGCTGCCCCAGAAAG 3816
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 3942 GTTCGACCAAGGCAAGATCTGCAAGTGAAGCCACCCAGAGTGTGATTCGAACRAGGG 4001
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; QY 3817 CATCTATTTTCCATGTCTAGTGTCCAGGCTGGAGCAAG 3856
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 4002 CCGGTATTAGAGTGTCTCAATCTGCAGAGTCTCGGAAAG 4041
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

; RESULT 13
; US-09-996-545-1
; Sequence 1, Application US/08996545
; Patent No. 592898
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene atrd of

```

```

; TITLE OF INVENTION: Aspergillus Nidulans
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,545
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11766
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-2763
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4002
; US-08-996-545-1

```

```

Query Match 14.2%; Score 606.8; DB 2; Length 4002;
Best Local Similarity 49.4%; Pred. NO. 1.2e-149;
Matches 1768; Conservative 0; Mismatches 1757; Indels 57; Gaps 5;

```

```

QY 334 CAACCATCTGGAGGAGAAATGACCACGATGCCATATTTATACAGTGGGATCGGTCTGG 393
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 CGACGAGTCTTATGATGAATGACCAAGAACGTACTGTACTTCGTATACCTCGGTATCGG 482
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 394 CGTGTGGTGGCTGTACATCCAGTTTCATTTCTGGTGCCTGGCAGCAGGAGACAGAT 453
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 CGAGTTTGTACAGTCTATGTTAGTACTGTTGGCTTCATCTATACGGAGACACGCCAC 542
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 454 ACTCAAAATAGAAAATATTTTTCATGCTATCATGCGACAGAGATTTGGCTGTTTGA 513
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 543 GCAGAAGATCCCGAGTATTAAGTCTATCTCTGCGCCAGAACATTTGGCTATTTGA 602
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 514 CGTGCATGAGCTTGGGAGCTTAACACCCGCTCACAGACGATGCTCCAAATCAATGA 573
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 603 TAACTCGGTGGCGGGAAGTGAACCCGCTATACAGCCGATACAGCCGATCAAACTTATCCAGA 662
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 574 AGGAATTCGCGACAAAATTTGGAATTTCTTTCACCTCAATAGCAACATTTTTCACCGGTTT 633
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 663 TGGCATTTTCGGAAGGTCGGTCTCACTTTGACTGCCCTGGCGACATTCGTGACAGCAT 722
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 634 TATAGTGGGGTTTACAGTGGTTTGAAGCTAACCCCTGTGATTTTGGCCATCAGCCCTGT 693
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 723 CATATCGCCTAGCTCAAAATACTGGAAGTTGGCTCTAATTTGACGCTCAACAATTTGTGTC 782
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 694 TCTTGGACTTTTTCAGCGCCATCTGGCAAGATATCTATCTTCTTCTTACTGATAAGAACT 753
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 783 CCTCGTTCTCACCATGGCGGTTTCTCAGTTTATCATCAAGTACAGCAAAAGTCTCGT 842
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 754 CTTGGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTAGCAGCAATCAGAACTGT 813
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 843 TGACAGCTACGGTGCAGCGGCACATGTTGCGGAAGAGGTCAATCAGCTCCACTCAGAAATGCG 902
QY 814 GATTGCCCTTTGGAGGACAAAAGAAGAACTTTGAAGGTTACAAACAAAATTTAGAAAGACG 873
Db 903 CACACGGTTTGGCACCAAGACAAAGCTTGCAGACGAGTATGAGGTCCACATTAGACGAGC 962
QY 874 TAAAGGAATTTGGGATAAAGAAAGCTATCACCGGCAACAACTTTCTATTTGGTGGCGCTTTCTT 933
Db 963 TGAGAAATTTGGGACAAAAGAACCAGATTTGTCATGGGTTTTCATGATTTGGCGGCATGTTTGG 1022
QY 934 ATTGATCTATGCATCATATGCTCTGGCTTTCTGGTATGAGGACCTTCTTGGTCTCTCCAG 993
Db 1023 CCTTATGACTCGAACTACGGTCTGGCTTCTGGATGGGTTCTCGTTTCTTGGTAGATGG 1082
QY 994 TGAATATTCTATTGGACAAAGTACTACACTGTCTTCTTTCTGTTATTAATTTGGGGCTTTTAG 1053
Db 1083 TGCAGTCGATGGGTGATATCTCACAGTTCTCATGGCCATCTTGATCGGATGTTTCTC 1142
QY 1054 TATTTGGACAGGATCCCCAAGCATTTGAAGCATTTTGGCAACCGCAAGAGAGAGCTTTATGA 1113
Db 1143 CTTTGGGAACGTTAGTCCAAATGCTCAAGCATTTTACAACACGCTGTGGCGCGCGCGCAAA 1202
QY 1114 AATCTTTCAAGATAATTGACAATAAACCAAGCATTTGACAGCTATTGCGAAGAGTGACATAA 1173
Db 1203 GATATTTTGGACGATCGATCGCCAGTCCCCATTTAGATCCATATTCGAAACGAAGGGAAGAC 1262
QY 1174 ACCAGATAATATTAAAGGAAATTTTGGAAATTCAAAATGTTTCACTTCAGTTTACCCCTTCTCG 1233
Db 1263 GCTCGACCATTTTGGAGGCCACATTTGAGTTAGCCAAATGTCGAAGCATTTTACCCTATCTAG 1322
QY 1234 AAAAGAACTTAAGATCTTAAAGGTTCTCAACCTGAAGGTTTCAGAGTTGCGGACAGACATGGC 1293
Db 1323 ACCCGAGGTCAACCGTCATGGAGGATGTTTCTCTGTCATATGCGCGCTGTGMAAAACAACCGC 1382
QY 1294 CTTGTTTGGACAGTGTGCTCGGGAAGACGACGACCGCTGCAGCTGATGCAGAGGCTCTA 1353
Db 1383 TTTAGTGGCCCCCTCTGGCTCTGGAAGGATACGGTGGTGGCTTGGTTGAGCGATTCTA 1442
QY 1354 TGACCCACAGATGTCATGCTGTGATTGATGGACAGGACATTAAGACCAATAAATGTAAG 1413
Db 1443 CATGCCCTTTCGCGGTACGGTTTGTCTGGATGGCCATGACATCAAGNCCATCATCTCCG 1502
QY 1414 GCATCTTCGGGAATTTACTGGTGGTGAGTCAAGAGGCTGTGTTGTTTGGCCACCAAGAT 1473
Db 1503 CTGGCTTCGCCAACAGATCTCTTGGTTAGCCAGGAGCCTTCTCTTTTGGCAGCAGGAT 1562
QY 1474 AGCTGAACAACATTCGCTATGCCC-----CGGAAATGT 1506
Db 1563 TTATGAAGATATTAGGCACGGTCTCATCGGCACAAAGTACAGAAATGAATCCGAGGATAA 1622
QY 1507 CACCATGGATGAGATTGAAAGAGCTGTTAAGGAAGCCAAATGCCATATGATTTTATCATGAA 1566
Db 1623 GGTCCGGGAACCTCATCGAAGACGGCGCAAAATGCGGAATGCTCATGACTTTATTACTGC 1682
QY 1567 ACTACCTAATAATTTGACACTCTGGTTGGAGAGAGAGGGCCCCAGCTGAGTGGTGACA 1626
Db 1683 CTTCCTCTGAAGGTTATGAGACCAATTTTGGGACGGTGGCTTTCTCTCTTCAGGTGGCCA 1742
QY 1627 GAAACAGAGATTCGCATTTGCTCGGGCCCTGGTTGCGAACCCCAAGATTTCTCTGCTGGA 1686
Db 1743 GAAACAGCGCATTTGAATCGCCCGTGGCGTTGTTAGTGACCCCAAAATTCCTGCTCTGGA 1802
QY 1687 TGAGGCAACGTCAGCTCTGACACTGAAAGTGAAGCAGTGGTTCCAGGTGGGCCCTGGATAA 1746
Db 1803 TGAAGCTACTTTCGGCCTTTGGACACAAAATCCGAAGGCGTGGTTCAACGACGTTTGGAGAG 1862
QY 1747 GGCCAGAAAAGCCGAGCTACCATTTGATGTGAATGCTCATCTGTTTGTCTACAGTTTCGTAATGC 1806
Db 1863 GGCAGCTGAAGGCGGAACACTACTATTGTGATCGCTCATCGCTTTTCCACGATCAAAACGGC 1922
QY 1807 CGATGCTCATCTGTTTGTGATGGAGTCATTTGTGGAGAAAGGAATCATGATGAAC 1866
Db 1923 GCACACATTTGTTGTTCTGGTCAATGCGCAAAATTTGCTGAACAGGAAGTCAACGATGAAT 1982

QY 1867 CATGAAAGAGAAGGGCATTTTACTTTCAAACTTGT-----CACAAATGCAGACAAG 1914
Db 1983 GGTGACCGCGGAGCGCTTATCGCAAACTTTGTGGAGGCTCAACGTATCAATGAACAGAA 2042
QY 1915 AGAAATGAAATTTAGTTAGAAATGCCACTGGTGAATCCAAAGTGAAGTATGACCTT 1974
Db 2043 GGAAGCTGAGCCCTTGGAGAGCCGACGCTGAGGATCTCAAGATTCAGATATTGCCAA 2102
QY 1975 GGAATGTCTCCAAAAGATTACGGTCCAGTTTAAATAAAAAGAAAGATCAACTCGCAGGAG 2034
Db 2103 AATCAAACTGCTCAAGCGCATCATCCGATCTCGACGGAACCCCAACCAATTCACCG 2162
QY 2035 TATPACATGCACCAACAGGCCAAGACAGAAAGCTTTGTACAAAAGAGAGCTTTGAATGAGAA 2094
Db 2163 CACGGGCACCCACAAGTCTGTTTCCAGCGGATTTCTTTTAAAGACCC-----CCGAAAC 2219
QY 2095 TGTACCTCCAGTTTCTTCTGGAGGATTTCTGAAGCTG-----AATCAACTCA 2142
Db 2220 AACTCCGAAATACTCATATATGGACGCTGCTCAAAATTTGTTGCTTCAACCGCCCTGA 2279
QY 2143 ATGGCCTTATTTTGGTTGTTATATTTGTCTATATAAAACGGAGGCTTCAACCCAGC 2202
Db 2280 AATCCCGTACATGCTCATCGGTTCTTCTCAGTGTAGCTGGTGGTGGCCACCCAC 2339
QY 2203 ATTTTCAATAATATTTTCAAGGATTAAGGGA---TCTTTTACCCGAGATGAGGATTCCTGA 2259
Db 2340 GCAAGCAGTGTATATGCTTAAAGCCATCAGCACACTCTCGCTCCCAAGATCACAATATAG 2399
QY 2260 AACAAACGACAGAAATAGTAACATGTTTCTGTATTTGTTTCTAGTCTTGGAAATTTTTC 2319
Db 2400 CAAGCTTCGACATGTCGGGATTTCTGGTCAATGATGTTCTTCGTGGTTGGTATCATTTCA 2459
QY 2320 TTTTATTACATTTTTCCTCCAGGGCTTCAATTTGGCAAGCTGGGAGATCCTCACTAA 2379
Db 2460 GTTTATACGCACTCAACCAATGGTCTGATTTGCCGTATGCTCGAGAGACTTTATTCG 2519
QY 2380 GCGGCTTCGATACATGTTTTCAGATCCATGCTGAGACAGGATGTCAGCTGGTTTGAATGA 2439
Db 2520 TCGCGGAGAGACACTGCTTTTCGGACGATACTCCGCTCAAGACATTTGCTTCTTTTGACAA 2579
QY 2440 CCTAAAACACACACTGGAGCATTTGACACCAGGCTTGCATGATGCGGCTCAAGTTAA 2499
Db 2580 GGAAGAGATAGACCGCGCTGTGACCTTTCTCTGTCCACCGAGACGAAGCATCTCTC 2639
QY 2500 AGGGCTATAGGTTCCAGGCTTCTGCTCATATACCCAGAATATAGCAAAATCTTTGGGACAGG 2559
Db 2640 CGGTGTTAGCGGTGACCTTAGGCACGATCTTGATGACCTCCACGACCTTAGGAGCGC 2699
QY 2560 CATTTATATACCTTAATCTATGGTTGGCAATTAACACTTTTACTTTAGCAATTTGTACC 2619
Db 2700 TATCATTTATGCGCTGGCGATTTGGTGGAAATTTGGCTTAGTTGTTATCTCGGTTGTGCC 2759
QY 2620 CATCATTGCAATAGCAGGAGTTGTTGAAATGAAATGTTGCTGGACAGCACTGAAAGA 2679
Db 2760 GGTCTCTCGGCATCGGGTTTCTACCGATTTCTATATGCTAGCCGAGTTTCAATCAGGCTC 2819
QY 2680 TAAAGAAGAGCTAGAAGAGAGCTGGGAAGATTTGCTACAGAAGGCCATCGAAAACCTTCGGAAC 2739
Db 2820 CAAGCTTGCTTATGAGGATCTGCAAACTTTGCTTCGAGGCTACATCGTCTATCGCAC 2879
QY 2740 TGTGTTTCTTTGACTCGGAGCAGAAAGTTTGAATACATGTTATGACAGAGTTTGCAGAT 2799
Db 2880 AGTTGCTCATTAACCCGGGAAAGGATGCTGGAGATTTTACCATTGCCAGCTTTGACGC 2939
QY 2800 ACCATACAAAACCTTTTGGGNAAGCACACATCTCGGGGCTCTCATTTTCTCATCCCA 2859
Db 2940 ACAAGCAGGACGACGATCTAATCTCTGTTGAGGTCATCCCTGTTATATGCGTCTGCGA 2999
QY 2860 GGCATGATGTTTCTTCTGCTGCTGGCTTTCGGGTTTGGTGGCTACTTCTGTTGGCAAA 2919
Db 3000 GGCACATGTTTCTTCTGCTGCTGGCTTTCGGCTTGGTACGAGGAGACACTTCTTGGTCA 3059

Db 783 CCTCGTCTCACCATGGCGGTGTTCTCAGTTTATCATCAAGTACAGCAAAAAGTCGCT 842
QY 754 CTTGGCCTATGCAAAAGCTGGAGAGTAGCTGGAAGAAGTCTTAGCAGCAATCAGAACTGT 813
Db 843 TGACAGCTACGGTGCAGGCGGCACTGTTGCGGAAGAGGTCACTACGCTCCATCAGAAATGC 902
QY 814 GATTGCTTTGAGGAGCAAAAGAAAGAACTTGAAGAGTACAAACAAAATTTAGAAAGAC 873
Db 903 CACAGCTTTGGACCAAGACAGCTGCGGAAGCAGTATGAGGTCCACTTAGAGAAAGC 962
QY 874 TAAAGGAATTTGGATAAAGAAAGCTATCACGGCCCAACATTTCTATTGGTGGCGCTTTCTT 933
Db 963 TGAGAAATGGGAACAAAGAACAGATGTCATGGGTTTCATGATTTGGCGCCATGTTGG 1022
QY 934 ATTGATCTATGATCATATATGCTCTGGCTTTCTGGTATGGGACCTCCTTGGTCTCTCCAG 993
Db 1023 CCTTATGTACTCGAATACGGCTTTGGCTTCTGGATGGGTCTCGTTCCTTGGGTAGATGG 1082
QY 994 TGAATATTCTATTGGACAAGTACTCACTGCTCTTCTTTCTGTTATTAATTGGGGCTTTTAG 1053
Db 1083 TGCAGTCGATGGGTGATTTCTCACAGTTCTCATGGCCATCTTGATCGGATCGTTCTC 1142
QY 1054 TATTGGACAGGCATCCCCAAGCATTTGAACCATTTGCAACGCAAGAGAGCAGCTTATGA 1113
Db 1143 CTTGGGAAGCTTTAGTCCAAATGCTCAAGCATTTACAAACGCTGTGGCGCGCGCCAAA 1202
QY 1114 AATCTTCAAGATTAATTGACATAAACCAAGCATTTGACAGCTATTGCAAGAGTGGACATAA 1173
Db 1203 GATATTGGAAACATCGATCGCCAGTCCCCATTTAGATCCATATCGAACGAAGGAAGAC 1262
QY 1174 ACCAGATATATTAAAGGGAATTTGGAATTCAAAATTTTCACTTTCAGTTTACCCCTTCG 1233
Db 1263 GCTCGACCATTTTGGGGCCACATTTAGTTACGCATGTAAGCATATTTACCCATCTAG 1322
QY 1234 AAAAGAGTTAAGATCTTAAAGGCTCTCAACCTGGAAGTTTACAGTGGGACAGAGTGGC 1293
Db 1323 ACCGAGGTCACCGTCATGAGGATGTTCTCTGTCAATGCCGCTGGAACCAACACGCG 1382
QY 1294 GCTGGTTGGAAACAGTGGCTCGGGAAGAGACGACGCGTGCAGCTGATCAGAGGCTCTA 1353
Db 1383 TTTAGTCGGCCCTCTGGCTCTGGAAGAAAGTAGCGTGGTGGCTTGGTTGAGCGATCTTA 1442
QY 1354 TGACCCACACAGATGGCATGCTGTATGATGGACAGGACATPAGGACCATAAATGTAAG 1413
Db 1443 CATGCTGTTTCGCGGTACGGTTTTGCTGGATGCCATGACATCAAGGACCTCAATCTCG 1502
QY 1414 GCATCTCGGGAATTTACTGGTGTGAGTACAGAGCCTGTGTTTGGCCACACGAT 1473
Db 1503 CTGGCTTCGCCAACAGATCTCTTTGGTTAGCCAGAGCCTGTTCTTTTGGCAGCAGAT 1562
QY 1474 AGCTGAAAACATTCGCTATGCGC-----GGCAAAATGT 1506
Db 1563 TTATAAGAAATTTAGGCACCGTCTCATCGGCACAAAGTACGAGAATGAATCCGAGATA 1622
QY 1507 CACATGGATGAGATTGGAAGAGCTGTTAAGGAAGCAATGCCATGATGATTTATCATGAA 1566
Db 1623 GGTCCGGAACTCATCGAGAAGCGGCAAAATGGGGAATGCTCATGACTTTTATTACTGC 1682
QY 1567 ACTACCTAATAATTTGACACTCTGGTTGGAGAGAGGGCCACGCTGAGTGGTGACA 1626
Db 1683 CTTGCCCTAGAGTTATGAGACCAATGTTGGGACGCGTGGCTTTCTCCTTCAGGTGGCCA 1742
QY 1627 GAAACAGAAATCGCCATGCTCGGCGCTGTTGCGCAACCCCAAGATTTCTTCTGCTGA 1686
Db 1743 GAAACAGGCATTCATTCGCCGCTGGCTGTTGATGACCCCAAAATCCTGCTCTGGA 1802
QY 1687 TGAGGCAACGCTACGCTCTGACACTGAAAGTGAAGCAGTGGTTCAGTGCGCCCTGGAATA 1746
Db 1803 TGAAGCTACTTCGGCTTTGGACACAAATCCGAAGGCGTGGTTCAAGCAGCTTTGGAGAG 1862
QY 1747 GGCAGAAAGGCCGAGTACCATGTTGATGCTCATCTGTTTCTACAGTTCGTTAATGC 1806
Db 1863 GGCAGCTGAAGGCCGCACTACTATTGTGATCGCTCTCGCCTTTCCACGATCAAAAGCGC 1922

QY 1807 CGATGTCATTGCTGTTTGTATGATGGAGTCAATGTGGAGAAAGAAATCATGATGAAT 1866
Db 1923 GCACAACTTGTGTTCTGTCAATGSCAAAATTTGCTGAACAAGGAATCCTCAGATGAAT 1982
QY 1867 CATGAAAGAGAGGGCAATTTACTTCAAACTGT-----CACATGAGACAAG 1914
Db 1983 GGTGACCGGGGCGCTTATCGCAAACTTTGTGGAGGCTCAACGTATCAATGAACAGAA 2042
QY 1915 AGGAAATGAAATTCAGTTAGAAATGCCACTGTGTGAATCCAAAAGTGAAGTATGATGCTT 1974
Db 2043 GGAAGCTGAGCCCTTGGAGACGCCGAGCTGAGGATCTCACGAATCCAGATATGSCAA 2102
QY 1975 GGAATGTCTCCAAAAGATTACGGGTCAGTTTAATAAAAAGAGATCAACTCGCAGGAG 2034
Db 2103 AATCAAACTGCGTCAAGCGCATCATCCGATCTCGACGGAAACCCACAAACCATTTGACCG 2162
QY 2035 TATACATGCAACCAAGGCCAAGACACAAAGCTTTGTTACAAAGAGGACATGAATGAGAA 2094
Db 2163 CACGGGCAACCCACAAAGTCTGTTTCCAGCGGATTTCTTTAAAGAGCC---CCCGAAGC 2219
QY 2095 TGTACCTCCAGTTTCTCTGGAGGATTCCTGAAGCTG-----AACTCAACTGA 2142
Db 2220 AACTCCGAATACTCATATTGAGCGCTGCTCAAAATTTGTTGCTTCAACCGCCCTGA 2279
QY 2143 ATGGCCTTATTTTGTGTTGTTATTTTGTGCTATTATTAACGGAGGCGCTGCACACG 2202
Db 2280 AATCCGTTACATGCTCATCGGTCTGTTCTCAGTGTAGCTGTTGTTGGTGGCCAAACCCAC 2339
QY 2203 ATTTTCAATAATATTTCACGATTTATAGGA---TCTTTACCGAGATGAGGATCTGA 2259
Db 2340 GCAAGCAGTGTATGCTAAAGCCACAGACACTCTCGCTCCCAAGATCACAATATAG 2399
QY 2260 AACAAACGACAGAAATAGTAACATGTTTCTGTATTGTTTCTAGTCTTGGAAATTTATTC 2319
Db 2400 CAAGCTTCGACATGTCGGGATTTCTGTCATTGATGTTCTTCTGTTGTTGATCATCA 2459
QY 2320 TTTTATTAATTTTCTCCAGGGCTTCAATTTGGCAAGCTGGGAGATCTCTCACTAA 2379
Db 2460 GTTTATACGAGCAACCAATGGTGTGCTATTTGCCGATGTCCTCCGAGAGACTTATTCG 2519
QY 2380 GCGGCTTCGATACATGTTTTCAGATCCATGCTGAGACAGGATGTACAGCTGTTTATGA 2439
Db 2520 TCGCCGAGAGACACTGCTTTCGACGATACTCCGTCACAGACATGCTTTCCTTTGACAA 2579
QY 2440 CCTTAAAAACACCTGGAGCAATTGACAAACAGGCTTCCCAATGATCGGGCTCAAGTTAA 2499
Db 2580 GGAAGAGAATAGCACCGCGCTCTGACCTCTTCTGTCACCGAGAGACAGCATCTCTC 2639
QY 2500 AGGGCTATPAGTTCCAGGCTTGTGTCATTTACCCAGAAATATAGCAATCTTGGACAGG 2559
Db 2640 CGGTGTTAGCGGTGTGACTCTAGGCACGATCTGTATGACCTCCAGACCCCTAGGAGCGGC 2699
QY 2560 CATTTATATATCCTTAATCTATGTTGGCAATTAACACTTTTACTCTTTPAGCAATTTGTA 2619
Db 2700 TATCATTTATGCGCTGGCGATTTGGTGGAAATTTGSCCTTAGTTGTTGATCTCGTTGTC 2759
QY 2620 CATCATTCGAATPAGCAGGATTTGAAATGAAATGTTGTTGTCGACAGACACTGAAGA 2679
Db 2760 GGTTCCTCCTGGCATCGGTTTCTACCGATTTCTATATGCTAGTCCCGAGTTTCAATCAGG 2819
QY 2680 TAAAGAGAGCTAGAAAGGCTGGGAAGTTCGTACAGAACCATTCGAAAACCTTCCGAAC 2739
Db 2820 CAAGCTTGCTTATGAGGATCTGCAAACTTTGCTTGGAGGCTACATCGTCTATCCGCAC 2879
QY 2740 TGTGTTTCTTGTGCTCGGGAGCAGAAAGTTTGAATACATGATGACAGAGTTTGCAGAGT 2799
Db 2880 AGTTGCTCATTAACCCGGAAAGGATGCTCTGGGAGATTTTACCATGCCAGCTTGACGC 2939
QY 2800 ACCATACAGAAACTCTTTGAGGAAAGCAGACATCTTTCGGGCTCTCATTTTCTATCACC 2859
Db 2940 ACAAGCAGGACCAAGTCTAATCTGCTTGTAGGTCTATCCCTGTTTATATGCTGCTCGCA 2999

```

* QY 2860 GCGAATGATGATTTTCCCTATGCTGGCTGTTCCGGTTTGGTGCCTACTTGTGGCAAA 2919
Db 3000 GGCACCTGTTTCTCTCGCTGGCTCGGGTTTGGTACGGAGGACACTTCTTGGTCA 3059
QY 2920 TGAGTTTCAAGACCTTCAGGATGTTCTTGGTATTCTCAGCTATTGCTTGTGGTCCAT 2979
Db 3060 CCAGAGATGACATTTTCCGCTCTTGTGTTCTCCGAGATTCTTGTGGTCTCA 3119
QY 2980 GGCAGTGGGCGAGTCACTTCTGCTGCTGACTATGCCAAGCCAAAGTATCAGCAGC 3039
Db 3120 ATCCGGCGGACCGCTCTTTCCTTTCACAGACATGGGCAAGCGGAAGATCGGCGGC 3179
QY 3040 CCAGCTCATCATGATCATGAAAGAACCCCTCTGATTGACAGCTACAGCCCTCACGGCT 3099
Db 3180 CGAATTCGAGCAGCTGTTCGACCGAAGCCACAAATTGATACTGGTCTGAAGAGGGCA 3239
QY 3100 CAAGCCAAATACCTTGAAGGAATGTGACATTTAATGAGGTCTGTTCAACTATCCAC 3159
Db 3240 GAAGCTCGAAGCGTGAAGGTGAATCGAATTTAGGAACGTGCACCTTCAGATACCCGAC 3299
QY 3160 TCAGCCAGACATCCCGCTGCTCCAGGGCTGAGCCTCGAGGTGAAGAGGGCCAGACCT 3219
Db 3300 CCGCCAGAACACCTGTCTCGCGGCTTGACCTGACCGTGAAGCCTGGACAAATATGT 3359
QY 3220 GGCCCTGCTAGTAGAGTGGCTGTGGGAAGAGCACAGTGTTCAGCTCTCTAGAGCGCTT 3279
Db 3360 TCGCTGTGTCGAGCCACCGGTTGTGCAAGAGTACCACCATTCGATTGCTTGAGCGCTT 3419
QY 3280 CTATGACCCCTTGGCTGGTTGCTAGTCTAATTTGATGGCAAGAGATTAAGCACCTGAATGT 3339
Db 3420 TTACGATCGATTGCGGGTCTACCTTGTGATGGGAAGGACATAAGTAACTAAATAT 3479
QY 3340 CCAGTGGCTCCAGCACACCTGGGATCGTCTCAGGAGCCCATCCTGTTGACTGCAG 3399
Db 3480 CAACCTCTTACCGACGTTTCTGCTACTGCTCAGCCAGGAGCGGACACTGTACCAGGCGAC 3539
QY 3400 CATTTGCCGAGAACATTTGCTATGAGACACAGCCGGTGTATCATCATGAAGAGATTAT 3459
Db 3540 CATCAAGGAACATCTTACTTGTGTTGTCGAAGATGACGTACCGGAAGATTCTTGAT 3599
QY 3460 CGAGCGCCCAAGGCGCAACATACACCTTCTCATGAGACACTCCCTCGAGAAATACAA 3519
Db 3600 TAAGGCTTGAAGGACGCTAATATCTACGACTTCATCATGCTCCCGAGGGCTTTAA 3659
QY 3520 CACAGAGTAGGACAGAAAGCCAGCTCTGTGGGCCAGAAACAGCGCATTTGCCAT 3579
Db 3660 TACAGTTGTTGGCAGCAAGGAGGAGCATGTTGCTGCGGCCAAAGCAAGCTGTGCCAT 3719
QY 3580 AGCTCGCGCTCTTGTAGACAGCCTCATATTTGCTTTTGGATGAAGCTACATCAGCTCT 3639
Db 3720 TGCCCGAGCCCTTCTCGGGATCCCAAAATCCTTCTTCGATGAAGCGACGTCAGCCCT 3779
QY 3640 GGATACAGAAAGTCAAAAGTTTCCAAAGCCCTGGACAAAGCCAGAGAGGCGCGAC 3699
Db 3780 CGACTCCGAGTCAAGAAAGTCTCCAGGCGGCTTGGATGCGCTGCCCGAGGCGGAC 3839
QY 3700 CTGATTTGATCGCCCGCCGCTTGTCCACCATCCAGATGAGATTTAATAGTGGTGT 3759
Db 3840 CACAATCGCGGTTGCACCGACTAGCACGATTCAAAAGGCGGAGCGTTATCTATGTTTT 3899
QY 3760 TCAGATGGCAAGTCAAGAGCATGGCACATCAACAGCTGCTGGCCCGAAGAGGAT 3819
Db 3900 CGACCAAGGCAAGATCGTGAAGCGGAACGCACAGCGAAGCTGGTCCAGAAAAGGGCG 3959
QY 3820 CTATTTTTCATGCTCAGTGTCCAGCTGGACCAAGCGCTA 3861
Db 3960 GTACTAGGAGTGGTCACTTGACAGCTTGGCAAGGGCCA 4001

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 17:45:54 ; Search time 31.4943 Seconds
(without alignments)
4517.824 Million cell updates/sec

Title: US-09-672-725C-23

Perfect score: 6474

Sequence: 1 MDPEGRKGSAEKNFWKMGK.....LLAQKGIYFSMVQAGAKR 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	6474	100.0	1281	22 AAE00308
2	6469	99.9	1281	22 AAE00303
3	6468	99.9	1281	22 AAE00309
4	6461	99.8	1281	22 AAE00310
5	6425.5	99.3	1280	22 AAB81068
6	6425.5	99.3	1280	22 AAE00304
7	5866	90.6	1280	18 AAW44073
8	5866	90.6	1280	21 AAY58186
9	5866	90.6	1280	22 AAB81959
10	5862	90.5	1280	22 AAB81064
11	5859.5	90.5	1283	22 AAB81065

12	5857	90.5	1280	21 AAY58187	Human G185V mutant
13	5857	90.5	1280	22 AAB81066	Human P-glycoprote
14	5857	90.5	1280	22 AAE00306	Human P-glycoprote
15	5854	90.4	1280	8 AAP70452	Sequence encoded b
16	5851	90.4	1280	11 AAR04868	Protein encoded by
17	5849.5	90.4	1279	22 AAB81067	Human P-glycoprote
18	5849.5	90.4	1279	22 AAE00307	Human P-glycoprote
19	5844	90.3	1280	15 AAR63624	Human P-glycoprote
20	5841	90.2	1280	14 AAR44297	Sequence encoded b
21	5841	90.2	1280	22 AAU04347	Human multidrug re
22	5840.5	90.2	1279	19 AAW80294	Mutated human P-gl
23	5830	90.1	1280	19 AAW48997	Wild-type human P
24	5824	90.0	1280	19 AAW48998	N-terminal single
25	5824	90.0	1280	19 AAW48999	C-terminal single
26	5818	89.9	1280	19 AAW49000	Human P glycoprote
27	5641	87.1	1276	21 AAY58189	Murine multidrug r
28	5621	86.8	1272	21 AAY70597	Rat multidrug resi
29	5611	86.7	1272	21 AAY70596	Rat multidrug resi
30	5301	81.9	1276	14 AAR35199	Mouse multidrug re
31	5301	81.9	1276	21 AAY58188	Murine multidrug r
32	5291.5	81.7	1275	21 AAY78879	Rat mdrlb2 (multis
33	5291.5	81.7	1275	22 AAB60409	Rat mdrlb2 multidr
34	2810.5	43.4	1275	20 AAW82594	H. contortus PGP-A
35	2693	41.6	1313	22 ABB63904	Drosophila melanog
36	2555	39.5	1279	22 ABB60234	Drosophila melanog
37	2512	38.8	1349	20 AAY16434	Multiple drug resi
38	2491.5	38.5	1302	22 ABB64919	Multiple drug resi
39	2482	38.3	1334	20 AAY02630	Drosophila melanog
40	2446	37.8	1408	19 AAG62871	Protein encoded by
41	2423	37.4	1286	21 AAG39101	Multiple drug resi
42	2398	37.0	1333	22 AAB62495	Arabidopsis thalia
43	2397	37.0	1243	21 AAG39102	Arabidopsis thalia
44	2383	36.8	1320	22 ABB64922	Drosophila melanog
45	2382	36.8	1254	21 AAY92173	A. thaliana ATPAC,

ALIGNMENTS

RESULT 1

AAE00308

ID AAE00308 standard; Protein; 1281 AA.

XX AC AAE00308;

XX DT 13-JUN-2001 (first entry)

XX DE Dog P-glycoprotein (PGP) allelic variant (Genotype A).

XX KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;

XX KW MDRI; drug bioavailability; transgenic animal; genetic model.

XX OS Canis familiaris.

XX FH Key Location/Qualifiers

XX FT Misc-difference 197 /note= "His of GenotypeC substituted by Gln"

XX FT WO200123540-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26767.

XX PR 28-SEP-1999; 99US-0156510.

XX PA (GENT-) GENTEST CORP.

XX PI Stocker PV, Stelmel-crespi DT, Crespi CL, Reif TC, Patten CJ;

XX DR WPI; 2001-235373/24.

XX DR N-PSDB; AAD03504.

XX

PT New dog P-glycoproteins (pgp) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
XX
XX Claim 17; Page 91-93; ilpp: English.
XX
CC The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDR1) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
CC This sequence is also referred as Genotype A protein. The pgp
CC enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. this enzyme is a member of the ABC
CC transporter family.
XX
SQ Sequence 1281 AA;
Query Match 100.0%; Score 6474; DB 22; Length 1281;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPGGRKGSAAENFWMKMGKSSKKEKKEKPTVSTFAMFYSNWLDRMLVGTMAAII 60
DB 1 MDPGGRKGSAAENFWMKMGKSSKKEKKEKPTVSTFAMFYSNWLDRMLVGTMAAII 60
QY 61 HGAALPLMLVFGNMTDSFANAGISRNKTFPPVIINESITNTQHFINHLEEMTYAYYY 120
DB 61 HGAALPLMLVFGNMTDSFANAGISRNKTFPPVIINESITNTQHFINHLEEMTYAYYY 120
QY 121 SGTGAGVLVAAYTOVSFWCLAAQRILKIRKQPFHAIMRQEGFWDVHDVDELNTRLTD 180
DB 121 SGTGAGVLVAAYTOVSFWCLAAQRILKIRKQPFHAIMRQEGFWDVHDVDELNTRLTD 180
QY 181 VSKINEGDKIGMFTQSIAFTFTGTVGTRGWKLTVLILASPVGLSAAIAWKILSS 240
DB 181 VSKINEGDKIGMFTQSIAFTFTGTVGTRGWKLTVLILASPVGLSAAIAWKILSS 240
QY 241 FTDKELLAYAKAGAAEVLAAIRTVIAGGQKKELEKERNKNLEAKGIGIKKAITANIS 300
DB 241 FTDKELLAYAKAGAAEVLAAIRTVIAGGQKKELEKERNKNLEAKGIGIKKAITANIS 300
QY 301 IGAFLLIYASALAFWYGTSLVLSSEYSIGQVLTFFSVLIGAFSIGQASPSIEAFANA 360
DB 301 IGAFLLIYASALAFWYGTSLVLSSEYSIGQVLTFFSVLIGAFSIGQASPSIEAFANA 360
QY 361 RGAAYETFKIIDNKPSIDTSYKSGHGPNDNTKGNLEKFNHVSFSPSRKEVKILGLNLKVQ 420
DB 361 RGAAYETFKIIDNKPSIDTSYKSGHGPNDNTKGNLEKFNHVSFSPSRKEVKILGLNLKVQ 420
QY 421 SGQTVLVGNSGCGKSTTVQMLQRLDPTDGMVCIDGQDRTINVRHLREITGVWSQEPV 480
DB 421 SGQTVLVGNSGCGKSTTVQMLQRLDPTDGMVCIDGQDRTINVRHLREITGVWSQEPV 480
QY 481 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFTDLVGERGAQLSGGQKOR 540
DB 481 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFTDLVGERGAQLSGGQKOR 540
QY 541 IATARALVRNPKILLDEATSDLTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
DB 541 IATARALVRNPKILLDEATSDLTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
QY 601 AGFDGQVIVKGNHDELMKEKGIYFKLVMTQTRGNETELEENATGESKESDALEMPKDS 660

DB 601 AGFDGQVIVKGNHDELMKEKGIYFKLVMTQTRGNETELEENATGESKESDALEMPKDS 660
QY 661 GSSILIKRRSTRRSIHAPQGDQRKLGKEDLNENVPVSFWRILKLNSTWEPYFVVGIFCA 720
DB 661 GSSILIKRRSTRRSIHAPQGDQRKLGKEDLNENVPVSFWRILKLNSTWEPYFVVGIFCA 720
QY 721 IINGLOPAPSIIPRSRIIGITRDEDEPETKRONSMFSLVLGLIISFTTFELQGFTEG 780
DB 721 IINGLOPAPSIIPRSRIIGITRDEDEPETKRONSMFSLVLGLIISFTTFELQGFTEG 780
QY 781 KAGETILKRLRYMVFERSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGAIKSLAVITQ 840
DB 781 KAGETILKRLRYMVFERSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGAIKSLAVITQ 840
QY 841 NIANLGTIIISLIYQWQLTLLLAIVPIIATAIGVEMKMLSGQALKDKKLEGGAKIAT 900
DB 841 NIANLGTIIISLIYQWQLTLLLAIVPIIATAIGVEMKMLSGQALKDKKLEGGAKIAT 900
QY 901 EAIENFRTVSLTREQKFEYMAQSLQVPPYNSLRKAHIFGVSFSTQAMMYFSYAGCFR 960
DB 901 EAIENFRTVSLTREQKFEYMAQSLQVPPYNSLRKAHIFGVSFSTQAMMYFSYAGCFR 960
QY 961 FGAYLVANEFMNFQDVLVLSAIVFGAMAGOVSSFPADYAKAKVSAAHVIMIEKSPLI 1020
DB 961 FGAYLVANEFMNFQDVLVLSAIVFGAMAGOVSSFPADYAKAKVSAAHVIMIEKSPLI 1020
QY 1021 DSYSPHGLKPNLTLEGNVTFNEVFNYTPRDPDPVQLGSLLEVKKGQTLALVGGSCGKST 1080
DB 1021 DSYSPHGLKPNLTLEGNVTFNEVFNYTPRDPDPVQLGSLLEVKKGQTLALVGGSCGKST 1080
QY 1081 VVOLLERYDPLAGSVLDGKEIKHLNVQWLRHGLTVSQEPILFDCSIAENIAYGNSR 1140
DB 1081 VVOLLERYDPLAGSVLDGKEIKHLNVQWLRHGLTVSQEPILFDCSIAENIAYGNSR 1140
QY 1141 VVSHEEIMQAAKEANIHHFIETLPEKYNTRYGDKGTQLSGQKQRIAIARALVRQPHILL 1200
DB 1141 VVSHEEIMQAAKEANIHHFIETLPEKYNTRYGDKGTQLSGQKQRIAIARALVRQPHILL 1200
QY 1201 LDEATSDLTSEKVVQALDKAREGRTCIIVIAHRLSTIQNADLIIVFQNGKVEHGTQ 1260
DB 1201 LDEATSDLTSEKVVQALDKAREGRTCIIVIAHRLSTIQNADLIIVFQNGKVEHGTQ 1260
QY 1261 QLLAQKGIYFSMVSVOAGAKR 1281
DB 1261 QLLAQKGIYFSMVSVOAGAKR 1281
RESULT 2
AAE00303
ID AAE00303 standard; Protein; 1281 AA.
XX
AC AAE00303;
XX
DT 13-JUN-2001 (first entry)
XX
DE Dog (PGP) P-glycoprotein (genotype C) #1.
XX
KW Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW drug bioavailability; transgenic animal; genetic model.
XX
OS Canis familiaris.
XX
PN WO200123540-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26767.
XX
PR 28-SEP-1999; 99US-0156510.
XX
PA (GENT-) GENTEST CORP.
XX
PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;

PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26767.
XX
PR 28-SEP-1999; 99US-0156510.
XX
PA (GENT-) GENWEST CORP.
XX
PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX
PI WPI; 2001-235373/24.
XX
DR N-PSDB; AAD03505.
XX
DR New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
PS Claim 17; Page 99-102; 11pp; English.
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
XX as multidrug transporter (MDR1) and nucleic acids encoding them.
XX The invention also includes fragments and biologically functional
XX variants of dog P-glycoprotein. PGP and their nucleic acids are
XX useful for determining the bioavailability of drugs and for
XX screening PGP inhibitors. They are useful for the diagnosis and
XX treatment of conditions characterised by PGP activity, by
XX reducing or increasing PGP activity in a cell. PGP nucleic acids
XX are used as oligonucleotide probes. Complements of PGP nucleic
XX acids are useful as antisense oligonucleotides, to induce a PGP
XX 'knockout' phenotype. They are used to prepare a non-human
XX transgenic animal, which are valuable as genetic models for
XX human diseases.
XX The present sequence is dog P-glycoprotein (PGP) allelic variant.
XX This sequence is also referred as Genotype B protein. The PGP
XX enzyme functions as an efflux pump exporting small molecules
XX across the cell membrane. This enzyme is a member of the ABC
XX transporter family.
XX
XX Sequence 1281 AA;
XX
Query Match 99.9%; Score 6468; DB 22; Length 1281;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1280; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDPEGRKGSAREKFNWKGKSKKKEKPPVSTFAMFRYSNWLDRMLVGTMAAII 60
DB 1 MDPEGRKGSAREKFNWKGKSKKKEKPPVSTFAMFRYSNWLDRMLVGTMAAII 60
QY 61 HGAALPLMLLVFGNMTDSFANAGISRKNTFPVIINESITNTQHFHNLHEEEMTTVAYYY 120
DB 61 HGAALPLMLLVFGNMTDSFANAGISRKNTFPVIINESITNTQHFHNLHEEEMTTVAYYY 120
QY 121 SGIGAGVLVAAYTQVSWFCLAAQRQLKIRKQFFHAIMRQEGWFDVHVGELNLTDD 180
DB 121 SGIGAGVLVAAYTQVSWFCLAAQRQLKIRKQFFHAIMRQEGWFDVHVGELNLTDD 180
QY 181 VSKINEGIGDKIGNFOSIAFTFTGTVGRGWKLTVLIALSPVLGSAALWAKILSS 240
DB 181 VSKINEGIGDKIGNFOSIAFTFTGTVGRGWKLTVLIALSPVLGSAALWAKILSS 240
QY 241 FTDKELLAYAKAGAVEVLAAITRTVAFGGQKKELEKYNKNLEAKGIGKAITANIS 300
DB 241 FTDKELLAYAKAGAVEVLAAITRTVAFGGQKKELEKYNKNLEAKGIGKAITANIS 300
QY 301 IGAFLILIYASALAFWYGTSLVLSSEYSIGQVLTFFSFLIGAFSIGQASPSIEAFANA 360
DB 301 IGAFLILIYASALAFWYGTSLVLSSEYSIGQVLTFFSFLIGAFSIGQASPSIEAFANA 360
QY 361 RGAAYEYFKIIDNKPSIDTSKSGHKPDNKNLEKFNHVSYPSPSRKVKILKGLNLKYQ 420
DB 361 RGAAYEYFKIIDNKPSIDTSKSGHKPDNKNLEKFNHVSYPSPSRKVKILKGLNLKYQ 420
QY 421 SGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCIDGQDRTINVRHLREITGVVSOEPV 480

DB 421 SGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCIDGQDRTINVRHLREITGVVSOEPV 480
QY 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGOKQR 540
DB 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGOKQR 540
QY 541 IAIARALVRNPKILLDEATSEALDTESEAVVQVALDKARKGRITTVIAHRLSTVRNADVI 600
DB 541 IAIARALVRNPKILLDEATSEALDTESEAVVQVALDKARKGRITTVIAHRLSTVRNADVI 600
QY 601 AGFDGVIIVEKGNHDELMKEGIYFKLVMTQTRGNETELENAETESKESDALEMPKDS 660
DB 601 AGFDGVIIVEKGNHDELMKEGIYFKLVMTQTRGNETELENAETESKESDALEMPKDS 660
QY 661 GSSLIKRSTRSTRSHAPQGDQRKLGTEDNENVPVPSFWRILKLNSTWEPYFVVGIFCA 720
DB 661 GSSLIKRSTRSTRSHAPQGDQRKLGTEDNENVPVPSFWRILKLNSTWEPYFVVGIFCA 720
QY 721 IINGLOPAFISIFRSIIIGITRDEDEPETKRONSMFVLVLGLGIISFTFFLQGTFFG 780
DB 721 IINGLOPAFISIFRSIIIGITRDEDEPETKRONSMFVLVLGLGIISFTFFLQGTFFG 780
QY 781 KAGEILTKRLRYMFRSMLRODVSFDDPKNTGALTTRLANDAAQVKGALSRNAVITQ 840
DB 781 KAGEILTKRLRYMFRSMLRODVSFDDPKNTGALTTRLANDAAQVKGALSRNAVITQ 840
QY 841 NIANLGTGIIISLYGWQLTLLLAIVPIIAIAGVEMKMLSGQALKDKKELEGAGKIAT 900
DB 841 NIANLGTGIIISLYGWQLTLLLAIVPIIAIAGVEMKMLSGQALKDKKELEGAGKIAT 900
QY 901 EAIENFRVTVSLTREQKFEYMAOSLOQVPYRNSLRKAHIFGVFSFISITQAMMYFSYACFR 960
DB 901 EAIENFRVTVSLTREQKFEYMAOSLOQVPYRNSLRKAHIFGVFSFISITQAMMYFSYACFR 960
QY 961 FGAYLVANEFNFQDVLVFSIAIVFGAMAVQVSSFAPDYAKAKVSAAHVIMIEKSPLI 1020
DB 961 FGAYLVANEFNFQDVLVFSIAIVFGAMAVQVSSFAPDYAKAKVSAAHVIMIEKSPLI 1020
QY 1021 DSYSPHGLKPNLTGNTVTFNEVFNYPTRDPVLPVGLGLEVKKQTLALVGGSCGCKST 1080
DB 1021 DSYSPHGLKPNLTGNTVTFNEVFNYPTRDPVLPVGLGLEVKKQTLALVGGSCGCKST 1080
QY 1081 VVQLLERYDPLAGSVLDGKEIKHLNVQWLRAHLGIVSQEPILFDCSIAENIAYGNSR 1140
DB 1081 VVQLLERYDPLAGSVLDGKEIKHLNVQWLRAHLGIVSQEPILFDCSIAENIAYGNSR 1140
QY 1141 VVSHEEIMQAAKEANIHHFIETLPEKYNTRYGDKGTQLSGGOKORIAIARALVRQPHILL 1200
DB 1141 VVSHEEIMQAAKEANIHHFIETLPEKYNTRYGDKGTQLSGGOKORIAIARALVRQPHILL 1200
QY 1201 LDEATSEALDTESEKVVQBALDKAREGRTCTIVIAHRLSTIQNALDIVVFQNGKVEHGTQ 1260
DB 1201 LDEATSEALDTESEKVVQBALDKAREGRTCTIVIAHRLSTIQNALDIVVFQNGKVEHGTQ 1260
QY 1261 QLLAOKGIYFMSVYQAGAKR 1281
DB 1261 QLLAOKGIYFMSVYQAGAKR 1281
RESULT 4
AAE00310
ID AAE00310 standard; Protein; 1281 AA.
XX
AC AAE00310;
XX
DT 13-JUN-2001 (first entry)
XX
DE Dog P-glycoprotein (PGP) allelic variant (Genotype D).
XX
KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
KW MDR1; drug bioavailability; transgenic animal; genetic model.
XX

OS Canis familiaris.

XX Key Location/Qualifiers
 FH Misc-difference 25 /note= "Asn of GenotypeC substituted by Lys"
 FT Misc-difference 197 /note= "His of GenotypeC substituted by Gln"
 FT Misc-difference 329 /note= "Ser of GenotypeC substituted by Thr"
 FT Misc-difference 1148 /note= "Met of GenotypeC substituted by Val"
 XX W0200123540-A2.
 XX 05-APR-2001.
 XX 28-SEP-2000; 2000WO-US26767.
 XX 28-SEP-1999; 99US-0156510.
 XX (GENT-) GENTEST CORP.
 XX Stocker PJ, Steinel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 XX WPI; 2001-235373/24.
 XX N-PSDB; AAD03506.

XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 PT for determining the bioavailability of drugs and for screening for dog
 XX PGP inhibitors -
 XX Claim 17; Page 108-110; 111pp; English.
 XX The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. Pgp and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype, they are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
 CC This sequence is also referred as Genotype D protein. The PGP
 CC enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.

XX Sequence 1281 AA;
 Query Match 99.8%; Score 6461; DB 22; Length 1281;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1278; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDPEGRGSAEKFNWKKSKKNEKKKPTVSTFAMFRYSNWLDRMLVGTWAAII 60
 DB 1 MDPEGRGSAEKFNWKKSKKNEKKKPTVSTFAMFRYSNWLDRMLVGTWAAII 60
 QY 61 HGAALPLMLVFGNNTDSFANAGISRNKTPFVINESITNNTQHFVNHLEEMTTAYYY 120
 DB 61 HGAALPLMLVFGNNTDSFANAGISRNKTPFVINESITNNTQHFVNHLEEMTTAYYY 120
 QY 121 SGIGAGVLVAAYIQVSWFCLAAAGQILKIRKQFFHATMRQEIFGWFDVHDVGNLRLTDD 180
 DB 121 SGIGAGVLVAAYIQVSWFCLAAAGQILKIRKQFFHATMRQEIFGWFDVHDVGNLRLTDD 180
 QY 181 VSKINEGIGDKIGMFQSIATFFTFGTVGRGKTLVLILAIISPVLGLSAAIWAKILSS 240
 DB 181 VSKINEGIGDKIGMFQSIATFFTFGTVGRGKTLVLILAIISPVLGLSAAIWAKILSS 240

QY 241 FTDKELLAYAKAGAVAEVLAARTVIAFGGOKKELERYNKNLEAKGIGIKKAITANIS 300
 DB 241 FTDKELLAYAKAGAVAEVLAARTVIAFGGOKKELERYNKNLEAKGIGIKKAITANIS 300
 QY 301 IGAAPLLIYASALAFWYGTSLVLSSEYSIGQVLTFFSVLIGAFSIGQASPSIEAFANA 360
 DB 301 IGAAPLLIYASALAFWYGTSLVLSSEYSIGQVLTFFSVLIGAFSIGQASPSIEAFANA 360
 QY 361 RGAAYEIFKIIDNKPSIDSYSGHKPDNLIKGNLEKFNHFSYPSRKEVKILKGLNLKQV 420
 DB 361 RGAAYEIFKIIDNKPSIDSYSGHKPDNLIKGNLEKFNHFSYPSRKEVKILKGLNLKQV 420
 QY 421 SGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDQIRTNVHRLREITGVVSOEPV 480
 DB 421 SGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDQIRTNVHRLREITGVVSOEPV 480
 QY 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAOLSGGQKOR 540
 DB 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAOLSGGQKOR 540
 QY 541 IAIARALVRNPKILLDDEATSDLTSEAVVOVALDKARKGRTTIVIAHRLSTVFNADVI 600
 DB 541 IAIARALVRNPKILLDDEATSDLTSEAVVOVALDKARKGRTTIVIAHRLSTVFNADVI 600
 QY 601 AGFDDGVIVEKGNHDELMKEKGIYFKLVTMTGRNEIELENATGESKSESDALEMSPKDS 660
 DB 601 AGFDDGVIVEKGNHDELMKEKGIYFKLVTMTGRNEIELENATGESKSESDALEMSPKDS 660
 QY 661 GSSLIKRRSTRSIIHAPOGQDRKLTGKEDLNENVPVSWFWRILKLNSTENWYFVVGIFCA 720
 DB 661 GSSLIKRRSTRSIIHAPOGQDRKLTGKEDLNENVPVSWFWRILKLNSTENWYFVVGIFCA 720
 QY 721 IINGGLQPAFSIIFSRIGIIFTRDEDPETKRONSNMFSVLFVLGIISFIFFLQGFTEG 780
 DB 721 IINGGLQPAFSIIFSRIGIIFTRDEDPETKRONSNMFSVLFVLGIISFIFFLQGFTEG 780
 QY 781 KAGEILTKRLRYWFRSMLRQDVSWFDDPKNTGALTTRLANDAAQVKGAGSLAVITQ 840
 DB 781 KAGEILTKRLRYWFRSMLRQDVSWFDDPKNTGALTTRLANDAAQVKGAGSLAVITQ 840
 QY 841 NIANLGTGIIISLIYGWLTLLLAIVPIIATAGVVENKMLSGQALDKKLELEGAGKIAT 900
 DB 841 NIANLGTGIIISLIYGWLTLLLAIVPIIATAGVVENKMLSGQALDKKLELEGAGKIAT 900
 QY 901 EAIENFRTVVSUTREQKFEYMYAOSLQVYPYRNSLRKAHIFGVFSITQAMMYFSYACFR 960
 DB 901 EAIENFRTVVSUTREQKFEYMYAOSLQVYPYRNSLRKAHIFGVFSITQAMMYFSYACFR 960
 QY 961 FGAYLVANEFMNFQDVLVFSNAIVFGAMAVQVSSFPADYAKAKVSAHVIMIEKSPLI 1020
 DB 961 FGAYLVANEFMNFQDVLVFSNAIVFGAMAVQVSSFPADYAKAKVSAHVIMIEKSPLI 1020
 QY 1021 DSYSPHGLKPNTLEGNTVFNEWFNYPTRPDIPIVQLGLSLEVKKGOTLALVSGSGCKST 1080
 DB 1021 DSYSPHGLKPNTLEGNTVFNEWFNYPTRPDIPIVQLGLSLEVKKGOTLALVSGSGCKST 1080
 QY 1081 VVOLLERYDPLASVLIDGKEIKHLNVQMLRAHLGTVSQEPILFDCSIAENIAYGDNRS 1140
 DB 1081 VVOLLERYDPLASVLIDGKEIKHLNVQMLRAHLGTVSQEPILFDCSIAENIAYGDNRS 1140
 QY 1141 VVSHEEIMOAAKEANIHHFIETLPEKYNTRVGDGTOLSGGQKORTAIARALVRQPHILL 1200
 DB 1141 VVSHEEIMOAAKEANIHHFIETLPEKYNTRVGDGTOLSGGQKORTAIARALVRQPHILL 1200
 QY 1201 LDEATSDLTSEKVVQEALDKAREGRTCTIVIAHRLSTIQNALDIIVVFONGKVKHEGTHQ 1260
 DB 1201 LDEATSDLTSEKVVQEALDKAREGRTCTIVIAHRLSTIQNALDIIVVFONGKVKHEGTHQ 1260
 QY 1261 OLLAQKGIYFSMVSVOQAKR 1281
 DB 1261 OLLAQKGIYFSMVSVOQAKR 1281

RESULT 5

AAB81068
ID AAB81068 standard; Protein; 1280 AA.

XX AAB81068;

XX 25-JUN-2001 (first entry)

XX Dog P-glycoprotein SEQ ID 7.

DE Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL;
KW efflux pump; dog.

XX Canis familiaris.

XX WO200123565-A1.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26592.

XX 28-SEP-1999; 99US-0156921.

XX 12-OCT-1999; 99US-0158818.

XX (GENT-) GENTEST CORP.

XX Stocker PJ, Steelmel-Crespi DT, Crespi CL;

XX WPI; 2001-316136/33.

XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell.

XX Claim 8; Page 73-76; 84pp; English.

XX This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDRL is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents
CC P-glycoprotein from a dog.

XX Sequence 1280 AA;

Query Match 99.3%; Score 6425.5; DB 22; Length 1280;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1273; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Qy 1 MDPEGRKGSAEKNFWKMGKSKKNEKKKPTVSTFAMFYSNWLRLYMLVGTMAAII 60

Db 1 MDPEGRKGSAEKNFWKMGKSKK- EKKEKAPTSTFAMFYSNWLRLYMLVGTMAAII 59

Qy 61 HGAALPLMWLVGNMTDSFANAGISRNNKTFPVIINESITNTQHFINHLEEMTTYAYY 120

Db 60 HGAALPLMWLVGNMTDSFANAGISRNNKTFPVIINESITNTQHFINHLEEMTTYAYY 119

Qy 121 SGIGAGVLVAAYIQVSFWCLAGROILKIRKOFFHAINRQIGFVDVHVGELNRLTDD 180

Db 120 SGIGAGVLVAAYIQVSFWCLAGROILKIRKOFFHAINRQIGFVDVHVGELNRLTDD 179

Qy 181 VSKINEGIDKIGMFFQSITATFTTGFIVGFTRGWKLTLVLAISPVLGLSAAIWAII 240

Db 180 VSKINEGIDKIGMFFQSITATFTTGFIVGFTPGWKLTLVLAISPVLGLSAAIWAII 239

241 FTDKELLAYAKAGAAVEEVLAAIIRTVIAFGGQKKELERNKNLEAKGIGIKKAITANIS 300
Db FTDKELLAYAKAGAAVEEVLAAIIRTVIAFGGQKKELERNKNLEAKRIGIKKAITANIS 299
Qy IGAFLLIYASYALAFWYGTSLVLSSEYSIGQVLTFFSVLIGAFSIGQASPSIEAFANA 360
Db IGAFLLIYASYALAFWYGTSLVLSSEYITIGQVLTFFSVLIGAFSIGQASPSIEAFANA 359
Qy RGAAYEIFKIIDNKPSIDSYSGHKPDNIKGNLEFNHVSYPSPKREVKILGLNLKVQ 420
Db RGAAYEIFKIIDNKPSIDSYSGHKPDNIKGNLEFNHVSYPSPKREVKILGLNLKVQ 419
Qy SGQTVLVGNSGGCKSTTVQIMQRLYDPTDGMVCIDQDRTINVRHLREITGVVSGEPV 480
Db SGQTVLVGNSGGCKSTTVQIMQRLYDPTDGMVCIDQDRTINVRHLREITGVVSGEPV 479
Qy LFATTIAENTRYGRENVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKOR 540
Db LFATTIAENTRYGRENVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKOR 539
Qy IAIARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
Db IAIARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 599
Qy AGFDDGVIVKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKDS 660
Db AGFDDGVIVKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKDS 659
Qy GSSLLKRESTRRSIHAPOGQRKLGTRKEDLNENVPVSFWRLKLNSTENWFFVVGIFCA 720
Db GSSLLKRESTRRSIHAPOGQRKLGTRKEDLNENVPVSFWRLKLNSTENWFFVVGIFCA 719
Qy IINGLQPAFSIIFSRIGIIFTRDDEPTKRONSMFSLVFLVGLIISFTIFFLQGGTFG 780
Db IINGLQPAFSIIFSRIGIIFTRDDEPTKRONSMFSLVFLVGLIISFTIFFLQGGTFG 779
Qy KAGEILTATRLRYWFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKAIGSLAVITQ 840
Db KAGEILTATRLRYWFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKAIGSLAVITQ 839
Qy NIANGTGIISLIYQWQTLTLLLAIVPIIAIAGVVENKMLSGQALKDKELEGAGKAT 900
Db NIANGTGIISLIYQWQTLTLLLAIVPIIAIAGVVENKMLSGQALKDKELEGAGKAT 899
Qy EAIENFRTVSLTRQKFEYMYAQSLQVPYRNSLRKAHIFGVFSFISITQAMVFSYAGCFR 960
Db EAIENFRTVSLTRQKFEYMYAQSLQVPYRNSLRKAHIFGVFSFISITQAMVFSYAGCFR 959
Qy FGAYLVANEFMFQDVLVLFSAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMIEKSPLI 1020
Db FGAYLVANEFMFQDVLVLFSAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMIEKSPLI 1019
Qy DSYSPHGLKPNLTLEGNVTFENVFNPTPRDIPVLQGLSLEVKKGOTLALVSGSGCKST 1080
Db DSYSPHGLKPNLTLEGNVTFENVFNPTPRDIPVLQGLSLEVKKGOTLALVSGSGCKST 1079
Qy VVOLLERFVDPYLAGSVLIDGKEIKHLNVQWLAHGLIVSQEPILFDCSTAEINAYGDNSR 1140
Db VVOLLERFVDPYLAGSVLIDGKEIKHLNVQWLAHGLIVSQEPILFDCSTAEINAYGDNSR 1139
Qy VVSHEETMAAKKANIHHFIETLPEKYNTVRGDKGQTLGGQKQRIATARALVRQPHILL 1200
Db VVSHEETMAAKKANIHHFIETLPEKYNTVRGDKGQTLGGQKQRIATARALVRQPHILL 1199
Qy LDEATSALDTESEKVVQVQALDKAREGRTICIVIAHRLSTIONADLIVVFONGVKKEGHTHQ 1260
Db LDEATSALDTESEKVVQVQALDKAREGRTICIVIAHRLSTIONADLIVVFONGVKKEGHTHQ 1259
Qy OLLAOKGIYFSMVSVQAGAKR 1281
Db OLLAOKGIYFSMVSVQAGAKR 1280

RESULT 6

AAE00304
XX AAE00304 standard; Protein; 1280 AA.
AC AAE00304;
XX
XX 13-JUN-2001 (first entry)
XX
XX Dog P-glycoprotein (PGP) #2.
XX
XX Dog; P-glycoprotein; PGP; multidrug transporter; MDRL;
KW drug bioavailability; transgenic animal; genetic model.
XX
XX Canis familiaris.
OS
XX WO200123540-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26767.
XX
XX 28-SEP-1999; 99US-0156510.
XX
XX (GENT-) GENTEST CORP.
XX
XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX
XX WPI; 2001-235373/24.
DR N-PSDB; AAD03489.
XX
XX
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
XX Example 2; Page 72-75; 111pp; English.
PS
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDRL) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variant of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP). The
CC PGP enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.
XX
XX Sequence 1280 AA;
SQ

Query Match 99.38; Score 6425.5; DB 22; Length 1280;
Best Local Similarity 99.48; Pred. No. 0;
Matches 1273; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
1;
QY 1 MDPEGGKGSAGKFWKMGKSKNEKKKPTVSTFAMFRYSNWLDRLYMLVGTWAAII 60
DB 1 MDPEGGKGSAGKFWKMGKSKK-ENKPKPTVSTFAMFRYSNWLDRLYMLVGTWAAII 59
QY 61 HGAALPLMLVFGNMTDSFANAGISRNKTPVINESITNNTQHFHILEEEMTTVAYYY 120
DB 60 HGAALPLMLVFGNMTDSFANAGISRNKTPVINESITNNTQHFHILEEEMTTVAYYY 119
QY 121 SGIGAGLVAAIYQVSWFCLAAAGROILKIRKQFFHAIMRQEGWFDVHDVGEINTRLTDD 180
DB 120 SGIGAGLVAAIYQVSWFCLAAAGROILKIRKQFFHAIMRQEGWFDVHDVGEINTRLTDD 179
QY 181 VSKINEGIGDKMGFFQSIAFTFTGFTVGRGKLTVLIAISPVLGSLAAIWAKILSS 240

DB 180 VSKINEGIGDKMGFFQSIAFTFTGFTVGRGKLTVLIAISPVLGSLAAIWAKILSS 239
QY 241 FTDKELLAYAKAGAAVEEVLAARTVIAFGQKKELERYKNLEAEKIGIKAITANIS 300
DB 240 FTDKELLAYAKAGAAVEEVLAARTVIAFGQKKELERYKNLEAEKIGIKAITANIS 299
QY 301 IGA AFLLIYASALAFWGTSLVLSSEYSIGOVLTVPFESVLICAFSIGQASPSIEAFANA 360
DB 300 IGA AFLLIYASALAFWGTSLVLSSEYTIQGVLTVPFESVLICAFSIGQASPSIEAFANA 359
QY 361 GAAAEIPIKIIDNKPSIDSYSGHKPDNIKGNLEFNHFSYPSRKEYKILKGLNLKQV 420
DB 360 GAAAEIPIKIIDNKPSIDSYSGHKPDNIKGNLEFNHFSYPSRKEYKILKGLNLKQV 419
QY 421 SGQTVLVGNSGGKSTTVOLMORLYDPTDGMVCIDQDITINVRHLREITGVVSOEPV 480
DB 420 SGQTVLVGNSGGKSTTVOLMORLYDPTDGMVCIDQDITINVRHLREITGVVSOEPV 479
QY 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGALSOGQKOR 540
DB 480 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGALSOGQKOR 539
QY 541 IATARALVRNPKILLIDDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
DB 540 IATARALVRNPKILLIDDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 599
QY 601 AGFDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELEENATGESKESDALEMSPKDS 660
DB 600 AGFDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELEENATGESKESDALEMSPKDS 659
QY 661 GSSLIKRRSTRRSIIHAPOGQDRKLGTEKDLNENVPVSWFWRILKLNSTWEPYFVVGIFCA 720
DB 660 GSSLIKRRSTRRSIIHAPOGQDRKLGTEKDLNENVPVSWFWRILKLNSTWEPYFVVGIFCA 719
QY 721 IINGGLQPAFISIFSRIGIFTRDEDPETHKRONSNWFSVLFLVGLIISFIITFFLQGTFFG 780
DB 720 IINGGLQPAFISIFSRIGIFTRDEDPETHKRONSNWFSVLFLVGLIISFIITFFLQGTFFG 779
QY 781 KAGEILTKRLYMWFRSMLRQDVSWFDDPKNTGALTTRLANDAAQVKGALGSRLAVITQ 840
DB 780 KAGEILTKRLYMWFRSMLRQDVSWFDDPKNTGALTTRLANDAAQVKGALGSRLAVITQ 839
QY 841 NIANGLTGIIISLYGWOLTLILLAIPIIAGVWEMKMLSGQALKDKKELEGAGKIAT 900
DB 840 NIANGLTGIIISLYGWOLTLILLAIPIIAGVWEMKMLSGQALKDKKELEGAGKIAT 899
QY 901 EAIENFRTVVSLTREQKFEYMAQSLQVVPYRNSLRKAHIFGVFSITQAMMYFSYACFR 960
DB 900 EAIENFRTVVSLTREQKFEYMAQSLQVVPYRNSLRKAHIFGVFSITQAMMYFSYACFR 959
QY 961 FGAYLVANEFMNFQDVLVFSALVFGAMAVQVSSFPADYAKAKVSAAHVIMIEKSPLI 1020
DB 960 FGAYLVANEFMNFQDVLVFSALVFGAMAVQVSSFPADYAKAKVSAAHVIMIEKSPLI 1019
QY 1021 DSYSPHGLKPNLTLEGNTVFNEWFNYPTRDPDIPVLOGLSLEVKKGTALVLSGGCGKST 1080
DB 1020 DSYSPHGLKPNLTLEGNTVFNEWFNYPTRDPDIPVLOGLSLEVKKGTALVLSGGCGKST 1079
QY 1081 VVOLLERYDPLAGSVLIDGKEIKHLNVQWLAHGLIVSQEPILFDCSIAENIAYGDSNR 1140
DB 1080 VVOLLERYDPLAGSVLIDGKEIKHLNVQWLAHGLIVSQEPILFDCSIAENIAYGDSNR 1139
QY 1141 VVSHEEIMQAANEIHHFETIPEKYNTRYGDKGTQSGQKQRIARALVRQPHILL 1200
DB 1140 VVSHEEIMQAANEIHHFETIPEKYNTRYGDKGTQSGQKQRIARALVRQPHILL 1199
QY 1201 LDEATSDALDTESEKVVQEALDKAREGRTCLVIAHRLSTQNADLIVVFONGVKEHGTHQ 1260
DB 1200 LDEATSDALDTESEKVVQEALDKAREGRTCLVIAHRLSTQNADLIVVFONGVKEHGTHQ 1259
QY 1261 QLLAQKGIIYFSWVSQAGAKR 1281
XXXXXXXXXXXXXXXXXXXX

Db 1198 LLDATSALEDESEKVVQVQALDKREGRTCTIVIAHRLSTIONADLIIVFVQNGRKEGTH 1257
 Qy 1260 QOLLAQKGIYFSMVSVQAGAKR 1281
 Db 1258 QOLLAQKGIYFSMVSVQAGTKR 1279

RESULT 8
 AAY58186
 ID AAY58186 standard; Protein; 1280 AA.
 XX
 AC AAY58186;
 XX
 DT 14-MAR-2000 (first entry)
 XX
 DE Human wild-type multidrug resistance-1 (MDR-1) protein.
 XX
 KW Multidrug resistance; MDR-1; P-glycoprotein;
 KW transmembrane efflux pump; haematopoietic stem cell; transduction;
 KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
 KW gene therapy; gene replacement; genetic defect; thalassaemia;
 KW Gaucher's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
 KW cytokine; wild-type.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 185
 FT /note= "This residue is Val in a mutant MDR-1
 FT (AAY58187)"
 XX
 PN WO9961589-A2.
 XX
 PD 02-DEC-1999.
 XX
 PF 27-MAY-1999; 99WO-US11825.
 XX
 PR 28-MAY-1998; 98US-0086988.
 XX
 PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 XX
 PI Sorrentino B, Bunting K;
 XX
 DR WPI; 2000-072615/06.
 DR N-PSDB; AAZ49332.
 XX
 PT Ex vivo expansion of hematopoietic stem cells transduced with a
 PT sequence encoding human multidrug resistance-1, used for bone marrow
 PT transplantation -
 XX
 PS Claim 10; Page 71-79; 113pp; English.
 XX
 CC This sequence represents human wild-type multidrug
 CC resistance protein MDR-1. MDR-1 is a transmembrane efflux
 CC pump, responsible for the export of drugs from certain
 CC cells, particularly cancer cells. Wild-type MDR-1 shows
 CC increased resistance to etoposide and decreased resistance to vinca
 CC alkaloids compared with a mutant form (AAY58187) where the Gly at
 CC position 185 is replaced by Val. The invention relates to transducing
 CC haematopoietic stem cells with nucleic acid encoding an MDR protein
 CC and culturing the modified cells. The modified haematopoietic stem
 CC cells are useful in bone marrow transplantation (to reconstitute
 CC haematopoietic systems in patients who have undergone chemotherapy or
 CC radiation therapy) and in ex vivo gene therapy of genetic defects in
 CC cells derived from haematopoietic stem cells, e.g., thalassaemia,
 CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
 CC cells can also be used to identify factors involved in regulating
 CC proliferation and differentiation in haematopoietic stem cells.
 CC Haematopoietic stem cells that express MDR-1 will be protected against
 CC chemotherapeutic agents, so can be engrafted while the patient is
 CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
 CC provides sufficient cells to permit standard biochemical analysis.

CC Overexpression of MDR-1 allows cytokine-driven expansion of
 CC haematopoietic stem cells by at least 10-fold compared with a maximum
 CC of 4-fold in known procedures.
 XX
 SQ Sequence 1280 AA;
 Query Match 90.6%; Score 5866; DB 21; Length 1280;
 Best Local Similarity 90.8%; Pred. No. 0;
 Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;
 QY 1 MDPEGGRKGS-A-EKNFWMKMKSSKKEKPKTPTVTAMPRYSNWLDRMLVGTMAAI 59
 DB 1 MDEGDRNGGAKKNFKLNKSEK-DKEKKPTVSFMSFYSNWLDRMLVGTMAAI 59
 QY 60 IHGAALPLMLVFGNMTDSFANAGISRNKTPFVINESITNNQTHFINHLEEMTYAY 119
 DB 60 IHGAGLPLMLVFGEMTDIFANAG-NLEDLMSNITNRSINDTGFEMN-LEEDMTRYAY 117
 QY 120 YSGIGAGVLVAAYIOVSFWCLAAAGROILKIRKQFPHAMRQEIFGWDVHDVDELNRLTD 179
 DB 118 YSGIGAGVLVAAYIOVSFWCLAAAGROILKIRKQFPHAMRQEIFGWDVHDVDELNRLTD 177
 QY 180 DVSKINEGDKIGMFFQSIATFFTFIVGTRGKLVILAIAPVLGLSAAIWAAIL 239
 DB 178 DVSKINEGDKIGMFFQSIATFFTFIVGTRGKLVILAIAPVLGLSAAIWAAIL 237
 QY 240 SFTDKELLAYAKAGAVEVLAARTVIAFGQKKELERYKNLEAKGIGIKKAITANI 299
 DB 238 SFTDKELLAYAKAGAVEVLAARTVIAFGQKKELERYKNLEAKGIGIKKAITANI 297
 QY 300 SIGAAFLIYASYALAFWGTSLVLSSEYSIGQVLTFFSVLIGAFSGQASPSIEAFAN 359
 DB 298 SIGAAFLIYASYALAFWGTSLVLSSEYSIGQVLTFFSVLIGAFSGQASPSIEAFAN 357
 QY 360 ARGAAEYIFKIIDNKPSIDSYKSGHKPDNIKGNLEKFNHVSFSPSRKEVKILKGLNKV 419
 DB 358 ARGAAEYIFKIIDNKPSIDSYKSGHKPDNIKGNLEKFNHVSFSPSRKEVKILKGLNKV 417
 QY 420 QSGQTVALVGNSSGCKSTTVQLMORLYDPTDGMVCIQODIRTNVRLREITGVVSEOP 479
 DB 418 QSGQTVALVGNSSGCKSTTVQLMORLYDPTDGMVCIQODIRTNVRLREITGVVSEOP 477
 QY 480 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDLTVGERGAQLSGGQK 539
 DB 478 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDLTVGERGAQLSGGQK 537
 QY 540 RIAIARALVRNPKILLDEATSALETSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 599
 DB 538 RIAIARALVRNPKILLDEATSALETSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 597
 QY 600 IAGFDDGVIVEKGNHDELMKEKGIYFKLVMTQTRGNEIELEENATGESKSDALEMSPKD 659
 DB 598 IAGFDDGVIVEKGNHDELMKEKGIYFKLVMTQTRGNEIELEENATGESKSDALEMSPKD 657
 QY 660 SGSLIKRRSRPRRIHAPOGQDRKLGKEDLNENVPVSWRILKLNSTWPFVVGIFC 719
 DB 658 SRSLIKRRSRPRSRVRSQADRLSKTEALDESIPVPSFWRIMKMLNLTWPFVVGIFC 717
 QY 720 AINGGLQPAFSIIFRSRIIGIFTRDEDPETKRONSMFSLFLVLGLIISITFFLQGF 779
 DB 718 AINGGLQPAFAIIFRSRIIGIFTRDEDPETKRONSMFSLFLVLGLIISITFFLQGF 777
 QY 780 GKAGEILTCLRVMFMSMLRQDVSFDDPKNTGALTTRLANDAAQVGAIGSLRAVIT 839
 DB 778 GKAGEILTCLRVMFMSMLRQDVSFDDPKNTGALTTRLANDAAQVGAIGSLRAVIT 837
 QY 840 QNTANLGTIIISLIYQWQLTLLLAIVPIAIVAGVVMKMLSGQALKDKELEGAGKTA 899
 DB 838 QNTANLGTIIISLIYQWQLTLLLAIVPIAIVAGVVMKMLSGQALKDKELEGAGKTA 897
 QY 900 TEAIENFRVTVSUTROKFFYMYAQSLQVPRNSLRKAHIFGVFSITQAMMYFSYAGCF 959
 DB 898 TEAIENFRVTVSUTROKFFYMYAQSLQVPRNSLRKAHIFGVFSITQAMMYFSYAGCF 957

QY 960 REFAYLVANEFNFODVLLVFSVAIVFGAMAVGVSSFPADYAKAKVSAHVIMIEKSPL 1019
DB 958 REFAYLVAHKLSFEDVLLVFSVAIVFGAMAVGVSSFPADYAKAKISAHHIIMIEKTPL 1017
QY 1020 IDSYSPHGLKPNLTGENTFENVEFNYPTRPDPVLOGLSLEVKKGTQALVGVSSGCGKS 1079
DB 1018 IDSYSTEGMLPNTLGNVTGFEVFNYPTRPDPVLOGLSLEVKKGTQALVGVSSGCGKS 1077
QY 1080 TVVQLLERYDPLAGSLVDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAENIAYGDSN 1139
DB 1078 TVVQLLERYDPLAGSLVDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAENIAYGDSN 1137
QY 1140 RVYSHEIMOAKENIHFITLPEKYNTRVGDGTOLSGQKORAIARALVROPHIL 1199
DB 1138 RVYSQEEIVRAAKEANIHAFTESLPNKYSTKVGDKGTOLSGQKORAIARALVROPHIL 1197
QY 1200 LDEATSAIDTESEKVVQALDKAREGRTCIIVIAHRLSTIQNADLIIVFQNGRVEKHGTH 1259
DB 1198 LDEATSAIDTESEKVVQALDKAREGRTCIIVIAHRLSTIQNADLIIVFQNGRVEKHGTH 1257
QY 1260 QOLLAQKGIYFMSVQAGAKR 1281
DB 1258 QOLLAQKGIYFMSVQAGTKR 1279

RESULT 9
AAB81959
ID AAB81959 standard; protein; 1280 AA.
XX
AC AAB81959;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human MDR1.
XX
KW Human; MDR1; multi-drug resistance; cholesterol uptake;
KW hypercholesterolaemia; hypcholesterolaemia; atherosclerosis;
KW coronary artery disease; cerebral vascular disease.
XX
OS Homo sapiens.
XX
PN WO200121762-A2.
XX
PD 29-MAR-2001.
XX
PF 22-SEP-2000; 2000WO-US26099.
XX
PR 23-SEP-1999; 99US-0155819.
XX
PA (BARN-) BARNES-JEWISH HOSPITAL.
PI Stenson WF, Tessner T;
XX
XX WPI; 2001-328100/34.
XX
XX Modulating cellular cholesterol uptake for treating hyper or
PT hypcholesterolemia, involves administering an agent that inhibits or
PT increases the expression of multiple drug resistance-1 -
XX
XX Example 1; Page 43-47; 47pp; English.
XX
XX The present invention describes a method of modulating cholesterol uptake
CC in cells, involving administering an agent capable of inhibiting the
CC multi-drug resistance protein MDR1. This is useful in the prevention and
CC treatment of cholesterol-related diseases, including
CC hypercholesterolaemia, hypcholesterolaemia, atherosclerosis, coronary
CC artery disease and cerebral vascular diseases. The present sequence is
CC the human MDR1 protein.
XX
SQ Sequence 1280 AA;
Query Match 90.6%; Score 5866; DB 22; Length 1280;

Best Local Similarity 90.8%; Pred. No. 0;
Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;
QY 1 MDPEGGKGS-A-EKNFWKMGKKKKNEKKKKPTVFTEAMFRYSNMLDRJLMLVGTMAAI 59
DB 1 MDLEGGDRNGAKKKNFKNKSEK-DKREKKPTVSVFMSFRYSNMLDKLVVVGTLAAI 59
QY 60 IHGAALPLMLLVFNGMTDFSPANAGISRNKTFPVIINESITNNQTHFHINLLEEMTYAYY 119
DB 60 IHGAGLPLMLLVFGEMTDIFANAG-NLEDLMSNITNRSINDTGFPMN-LEEDMTRYAYY 117
QY 120 YSGIAGVLVAAYIQVSWCLAAGRQLIKRKQFFHAIMRQEIQWEDVDHVGELNRLTD 179
DB 118 YSGIAGVLVAAYIQVSWCLAAGRQHKRKQFFHAIMRQEIQWEDVDHVGELNRLTD 177
QY 180 DVSKINGIGDKIMGFQSIATFTGFIQVGTGRGKLTLLVILAIISPVLGLSAAIAWAKILS 239
DB 178 DVSKINGIGDKIMGFQSIATFTGFIQVGTGRGKLTLLVILAIISPVLGLSAAIAWAKILS 237
QY 240 SFTDKELLAYAKAGAAVEVLAAIIRTVIAGGQKKELERYNKNLEAKGIGIKKAITANI 299
DB 238 SFTDKELLAYAKAGAAVEVLAAIIRTVIAGGQKKELERYNKNLEAKGIGIKKAITANI 297
QY 300 SIGAFLIIVASYALAFWYGTSLVLSSEYSTIGQVLTIVFFSVLIGAFSGQASPSIEAFAN 359
DB 298 SIGAFLIIVASYALAFWYGTSLVLSSEYSTIGQVLTIVFFSVLIGAFSGQASPSIEAFAN 357
QY 360 ARGAAEYEFKIIDNKPSIDSYSKSHKPDNIKGNLEKFNHFSYPSRKEYKILGLMLKV 419
DB 358 ARGAAEYEFKIIDNKPSIDSYSKSHKPDNIKGNLEKFNHFSYPSRKEYKILGLMLKV 417
QY 420 QSGQTVALVNGSGCKSTTVQLMORLYDPTDGMVCIQDQDITINVRHLREITGVWSQEP 479
DB 418 QSGQTVALVNGSGCKSTTVQLMORLYDPTDGMVCIQDQDITINVRHLREITGVWSQEP 477
QY 480 VLFATTAENIRYGRNVTMDIEIKAVKEANAYDFIMKLPNKFTLVGERGAQLSGGQK 539
DB 478 VLFATTAENIRYGRNVTMDIEIKAVKEANAYDFIMKLPNKFTLVGERGAQLSGGQK 537
QY 540 RIATARALVRNPKILLDEATSAIDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 599
DB 538 RIATARALVRNPKILLDEATSAIDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 597
QY 600 TAGFDDGVIVKEGNHDELMKEGIYFKLVTMTQTRNETELENAATGESKSDALEMSPKD 659
DB 598 TAGFDDGVIVKEGNHDELMKEGIYFKLVTMTQTRNETELENAATGESKSDALEMSPKD 657
QY 660 SGSSLIKRSTRRSIHAPQGDRLGTGKEDLNENVPVSEWRILKLNSTEMPYFVVGIFC 719
DB 658 SSSLIKRSTRRSVRGSOADRLSTKEALDESIPVPSFWRIMKLNLTENPYPVVGIFC 717
QY 720 AINGGLQPAFSIIPRIIGITRDEDEPETKRONSMESVLFVLGIIISFTFFLQGTFF 779
DB 718 AINGGLQPAFAIIFSKIIGVTRIDDEPETKRONSNLFSLLFLGALGIIISFTFFLQGTFF 777
QY 780 GKAGEILTKRLRYMVRSMRLQDVSWFDDPKNTTTCALTTLRLANDAAQVKGIGSLAVIT 839
DB 778 GKAGEILTKRLRYMVRSMRLQDVSWFDDPKNTTTCALTTLRLANDAAQVKGIGSLAVIT 837
QY 840 QNIANLGTGIIISLYIGWQLTLLLAIVPIIAIAGVEMKMLSGOALKDKKELEGAKIA 899
DB 838 QNIANLGTGIIISLYIGWQLTLLLAIVPIIAIAGVEMKMLSGOALKDKKELEGAKIA 897
QY 900 TEAIENFRTVSVLTREQFEYMYAQSLQVPYNSRKHAHIFGVSPSITQAMMYSIACGF 959
DB 898 TEAIENFRTVSVLTREQFEYMYAQSLQVPYNSRKHAHIFGVSPSITQAMMYSIACGF 957
QY 960 RFQAVLVANEENFODVLLVFSVAIVFGAMAVGVSSFPADYAKAKVSAHVIMIEKSPL 1019
DB 958 RFQAVLVANHKLMSFEDVLLVFSVAIVFGAMAVGVSSFPADYAKAKISAHHIIMIEKTPL 1017
QY 1020 IDSYSPHGLKPNLTGENTFENVEFNYPTRPDPVLOGLSLEVKKGTQALVGVSSGCGKS 1079

Db 1018 IDSYSTEGMNTLEGNNVTFGEVNFYTRPDIPVLOGLSLEVKKGOTLALVSGSGGKS 1077

Qy 1080 TVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRRAHLGIVSOEPTLFDCAENIAYGDNS 1139

Db 1078 TVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRRAHLGIVSOEPTLFDCAENIAYGDNS 1137

Qy 1140 RVSHHEIMQAQKANIHHFJETLPEKYNTRVGDGTQLSGGQKORIAIARALVRQPHIL 1199

Db 1138 RVSQEIEIVRAAKEANITHAFIESLPNKYSTKVGDKGTQLSGGQKORIAIARALVRQPHIL 1197

Qy 1200 LLDKATSLDTESEKVVQVEALDKAREGRTICIVIAHRLSTIQNADLIIVFONGKVKHEGTH 1259

Db 1198 LLDKATSLDTESEKVVQVEALDKAREGRTICIVIAHRLSTIQNADLIIVFONGKVKHEGTH 1257

Qy 1260 QOLLAQRGIYFSMYSVOAGAKR 1281

Db 1258 QOLLAQRGIYFSMYSVOAGTKR 1279

RESULT 10

ID AAB81064

XX AAB81064 standard; Protein; 1280 AA.

AC AAB81064;

XX 25-JUN-2001 (first entry)

XX Cynomologous monkey P-glycoprotein variant 1.

DE Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;

XX Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;

KW efflux pump.

XX Macaca fascicularis.

OS W0200123565-A1.

PN 05-APR-2001.

PD 28-SEP-2000; 2000WO-US26592.

PF 28-SEP-1999; 99US-0156921.

PR 12-OCT-1999; 99US-0158818.

XX (GENT-) GENTEST CORP.

XX Stocker PJ, Steimel-Crespi DT, Crespi CL;

PI WPI; 2001-316136/33.

DR N-PSDB; AAF86127.

XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein (PGP) and homologous PGP polypeptides are useful for predicting bioavailability of compound and increasing Pgp transporter activity in cell -

XX Claim 9; Page 57-59; 84pp; English.

XX This invention relates to a polynucleotide sequence encoding a cynomologous monkey P-glycoprotein (Pgp), and an allelic variant of the PGP protein. Pgp, also known as multidrug transporter, MDR1 is a member of the ABC transporter superfamily. The enzyme serves as an efflux pump exporting small molecules across the cell membrane. The invention includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence and protein, and also that of an allelic variant. The PGP polynucleotide sequence is useful for increasing PGP transporter activity in a cell. Antisense sequences of the cDNA are useful for inhibiting Pgp transport activity in a mammalian cell. They may also be used for increasing the bioavailability of a drug. The present sequence represents the cynomologous monkey P-glycoprotein.

XX Sequence 1280 AA;

SQ Query Match 90.5%; Score 5862; DB 22; Length 1280;

Best Local Similarity 90.7%; Pred. No. 0;

Matches 1163; Conservative 59; Mismatches 56; Indels 4; Gaps 4;

Qy 1 MDPEGGKGSAAE-KNFWMKGGKSKKNEKKEKPTVTFAMFRYSNWLDRYLMVGTMAAI 59

Db 1 MDLEGDRNGGAEEKNFKLNKSKK-DKKEKPTVSFMSFRYSNWLDRYLMVGTMAAI 59

Qy 60 IHGAALPLMLVGNMTDSFANAGISRNKTFPVIINESITNTQHFHINLEEMTYIAY 119

Db 60 IHGAGLPMLVFGMDTDFANAG-NLGDGLALLTNSSNITDTVPVYN-LEEDMTRYAY 117

Qy 120 YSIGAGVLVAAYIQVSWCLAAAGROILKIRQFPHAIMRQEIQFVDFVHVGELNRLTD 179

Db 118 YSIGAGVLVAAYIQVSWCLAAAGROILKIRQFPHAIMRQEIQFVDFVHVGELNRLTD 177

Qy 180 DVSKINEGIGDKIMGFFQSIATFTFTGFIYGFTRGKLTILVLAISPVGLSAAIAWAKILS 239

Db 178 DVSKINEGIGDKIMGFFQSMATFTFTGFIYGFTRGKLTILVLAISPVGLSAAIAWAKILS 237

Qy 240 SFTDKELLAYAKAGAAVEVLAIAIRTVIAFGQKKELERYKNLEAKGIGIKKAITANI 299

Db 238 SFTDKELLAYAKAGAAVEVLAIAIRTVIAFGQKKELERYKNLEAKGIGIKKAITANI 297

Qy 300 SIGAAFLLIYASVALAFWYGTSLVLSSEYSIGQVLTVPFSVLIGAFSGOASPSIEAFAN 359

Db 298 SIGAAFLLIYASVALAFWYGTSLVLSKEYSIGQVLTVPFSVLIGAFSGOASPSIEAFAN 357

Qy 360 ARGAAYEIFKIIDNKPSIDSYSGHKKPDNIKGNLEFNHFSYPSRKEVKILKGLNLKV 419

Db 358 ARGAAYEIFKIIDNKPSIDSYSGHKKPDNIKGNLEFNHFSYPSRKEVKILKGLNLKV 417

Qy 420 QSGQIVALVNGSGCGKSTTVQLMQLYDPTDMQVCIQDQDITINVRHLREITGVVSOEP 479

Db 418 QSGQIVALVNGSGCGKSTTVQLMQLYDPTDMQVSVQDQDITINVRHLREITGVVSOEP 477

Qy 480 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKPNKFDTLVGERGAQLSGGOK 539

Db 478 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKPNKFDTLVGERGAQLSGGOK 537

Qy 540 RIATARALVRNPKIILLDEATSALDTESEAVVQVVALDKARKGRTTIVIAHRLSTVRNADV 599

Db 538 RIATARALVRNPKIILLDEATSALDTESEAVVQVVALDKARKGRTTIVIAHRLSTVRNADV 597

Qy 600 IAGFDDGVIVEKGNHDELMKEGIYFKLVTMOTRONEIELENAATESSESDALEMSPKD 659

Db 598 IAGFDDGVIVEKGNHDELMKEGIYFKLVTMOTRONEIELENAATESSESIDTLEMSSHD 657

Qy 660 SGSSLIKRRSTRRSIHAPOGODRKLGTEDLNENVPVPSFWRILKLNSTWPFVVGIFC 719

Db 658 SGSSLIKRRSTRRSYRSGGQDRKLSTKEALDESIPPVSWFRIMKLNUTWPFVVGIFC 717

Qy 720 AIINGLOPAPSIIFRSIIGTIRDEDEPETKRONSMFSLVLVLIIGISFTIFFLQGTFF 779

Db 718 AIINGLOPAPFAVIPSIIIGTIRNDAETKRONSLFSLFLVLVLIIGISFTIFFLQGTFF 777

Qy 780 CKAGEILTKRLRYMFRSMRLRQDYSWFDPPKNTTGALTTRLANDAAQVKGAGISRLAVIT 839

Db 778 CKAGEILTKRLRYMFRSMRLRQDYSWFDPPKNTTGALTTRLANDAAQVKGAGISRLAVIT 837

Qy 840 QNIANLGTGIISLIYGHQLTLLLAIVPIIAIAGVEMKMLSGOALKDKKLEGGAKIA 899

Db 838 QNIANLGTGIISLIYGHQLTLLLAIVPIIAIAGVEMKMLSGOALKDKKLEGGAKIA 897

Qy 900 TEATENFTVYSLTREQKFEYNYAQSLQVPYRNSLRKAHIFGVGSFSTQAMMYSYAGCF 959

Db 898 TEATENFTVYSLTREQKFEYNYAQSLQVPYRNSLRKAHIFGVGSFSTQAMMYSYAGCF 957

Qy 960 RFGAYLVANEFMNTQDVLVLPFAIVFGAMAVGVSSFPADYAKAKVSAAHVMIIEKSP 1019

Db 958 RFGAYLVANEFMNTQDVLVLPFAIVFGAMAVGVSSFPADYAKAKVSAAHVMIIEKSP 1017

Qy 1020 IDSYSPHGLKENTLEGNNVTFGEVNFYTRPDIPVLOGLSLEVKKGOTLALVSGSGGKS 1079

Db 1018 IDSSTEGKLKPNLTBGNVTNEVFNYPTRLIDIPVQLGLSLEVKKGQTLALVSGSGGCKS 1077

QY 1080 TVVQLLERYDPLAGSLVDGKEIKHLNVQWLAHGLGIVSQEPILFDCSIAENIAYGDN 1139

Db 1078 TVVQLLERYDPLAGSLVDGKEIKQLNVQWLAHGLGIVSQEPILFDCSISENIAYGDN 1137

QY 1140 RVYSHEEIMOAANEIHHFIETLPEKYNTRVGDGKTQLSGGQKORIAIARALVRQPHIL 1199

Db 1138 RVYSQEEIVRAAKENIAHFIETLPEKYNTRVGDGKTQLSGGQKORIAIARALVRQPHIL 1197

QY 1200 LLDEATSALDTESEKVVQDALKAREGRTCVIAHRLSTIONADLIVVFQNGKVEHETH 1259

Db 1198 LLDEATSALDTESEKVVQDALKAREGRTCVIAHRLSTIONADLIVVFQNGKVEHETH 1257

QY 1260 QQLLAQKGIYFSMVSQAGAKR 1281

Db 1258 QQLLAQKGIYFSMVSQAGAKR 1279

RESULT 11

AAB81065

ID AAB81065 standard; Protein; 1283 AA.

XX

AC AAB81065;

XX

25-JUN-2001 (first entry)

XX

DE Cynomologous monkey P-glycoprotein variant 2.

XX

KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL1; efflux pump.

KW

XX

OS Macaca fascicularis.

XX

XX

FH Key Location/Qualifiers

FT Misc-difference 93..95

FT /note- "An additional 3 amino acids are present compared to PGP variant AAB81064"

FT

XX

PN WO200123565-A1.

XX

XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-US26592.

XX

PR 28-SEP-1999; 99US-0156921.

PR

PR 12-OCT-1999; 99US-0159818.

XX

XX

PA (GENT-) GENTEST CORP.

XX

XX

XX

PI Stocker PJ, Steelmel-Crespi DT, Crespi CL;

XX

DR WPI; 2001-316136/33.

DR

DR N-PSDB; AAF96128.

XX

XX

XX

PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein (PGP) and homologous PGP polypeptides are useful for predicting bioavailability of compound and increasing PGP transporter activity in cell

PT

PT

XX

XX

PS Claim 9; Page 65-68; 84pp; English.

XX

XX

CC This invention relates to a polynucleotide sequence encoding a cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the PGP protein. PGP, also known as multidrug transporter, MDRL1 is a member of the ABC transporter superfamily. The enzyme serves as an efflux pump exporting small molecules across the cell membrane. The invention includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence and protein, and also that of an allelic variant. The PGP polynucleotide sequence is useful for increasing PGP transporter activity in a cell. Antisense sequences of the cDNA are useful for inhibiting PGP transport activity in a mammalian cell. They may also be used for increasing the bioavailability of a drug. The present sequence represents the

CC cynomologous monkey P-glycoprotein variant 2. The protein has an additional 3 amino acids when compared to PGP variant 1 (AAB81065).

XX

QY

Sequence 1283 AA;

Query Match 90.5%; Score 5859.5; DB 22; Length 1283;

Best Local Similarity 90.4%; Pred. No. 0;

Matches 1165; Conservative 57; Mismatches 52; Indels 15; Gaps 5;

QY 1 MDPPGGKSGSAE-KNFWKMGKSKKNEKKEKPTVSTFAMFRYSNMLDLRLMLVGTMAAI 59

Db 1 MDLEBDRNGGAENKFFLANKSKK-DKEREKPTVSVFMSFRYSNMLDKLVMVGTAAI 59

QY 60 IHGAALPLMLLVFGNMTDSPANAG-----ISRNTKTPVVIINESITNTQTHINLLEE 112

Db 60 IHGAGLPLMLLVFGDMTDTFANAGNLGDLALLFNNT-----NSSNITDFVPVMN-LEED 113

QY 113 MTTAYYYSGIGAGLVAAAYIQVSWFCLAGROILKIRKOFFHAIMROEIGWDFVDHVG 172

Db 114 MTRYAYYSIGAGLVAAAYIQVSWFCLAGROILKIRKOFFHAIMRQEIIGWDFVDHVG 173

QY 173 LNRLTDDVSKINBEGIDKIGMPFQSIATFTGTFTGTRGWKLTLLVILASPVGLSAA 232

Db 174 LNRLTDDVSKINBEGIDKIGMPFQSMATFTGTGTRGWKLTLLVILASPVGLSAA 233

QY 233 IWAKILSFTDKELLAYAKAGAAVEVLAAIRTVIAFGGQKKELERYKNKLEAKGIGIK 292

Db 234 VWAKILSFTDKELLAYAKAGAAVEVLAAIRTVIAFGGQKKELERYKNKLEAKRIGIK 293

QY 293 KAITANTISGAFLIIVASVALAFWYGTSLVLSSEYSIGOVLTIVFFSVLIGAFSIGQASP 352

Db 294 KAITANTISGAFLIIVASVALAFWYGTSLVLSSEYSIGOVLTIVFFSVLIGAFSIGQASP 353

QY 353 STEAFANARGAAEYFIKIIDNKPSIDSYSGHGXPDNKGKLEKFNHVFSPSRKEVKIL 412

Db 354 STEAFANARGAAEYFIKIIDNKPSIDSYSGHGXPDNKGKLEKFNHVFSPSRKEVKIL 413

QY 413 KGLNLKVQSGQTVLAVNGSGCGKSTTVQLMORLYDPTDGMVCIDGODIRTNVHRLREIT 472

Db 414 KGLNLKVQSGQTVLAVNGSGCGKSTTVQLMORLYDPTDGMVSDGQDRTINVRFLREII 473

QY 473 GVSQEPVLFTATTAEINIRYRENVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQ 532

Db 474 GVSQEPVLFTATTAEINIRYRENVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQ 533

QY 533 LSGGQKORIAIARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLS 592

Db 534 LSGGQKORIAIARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLS 593

QY 593 TVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMOTRGTNELENADESSESDA 652

Db 594 TVRNADVIAGFDDGVIVEKGNHDELMKEKGIYFKLVTMOTRGTNELENADESSEIDT 653

QY 653 LEMSPKDSGSLIKRRSTRRSHAPQGDQRKLGTKEDLNENVPVPSFWRILKLNSTEMPY 712

Db 654 LEMSHDSSGSLIKRRSTRRSHAPQGDQRKLSTKEALDESIPPVPSFWRILKLNSTEMPY 713

QY 713 FVVGIFCAIINGGLOPAFSAIIFSRITIGITFDEDEPETKQNSNMFSLFLVLGIISITFF 772

Db 714 FVVGIFCAIINGGLOPAFAVIFSKIIGITFRNDDAETKQNSNLFSLFLVLGIIVSITFF 773

QY 773 FLOGFTFGKAGFELTKRLRYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAOVKGAIG 832

Db 774 FLOGFTFGKAGFELTKRLRYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAOVKGAIG 833

QY 833 SRLAVITONIANLGTGIIISLIYGNQLTLLLAIVPIITAIAGVVMKMLSGOALKDKKEL 892

Db 834 SRLAVITONIANLGTGIIISLIYGNQLTLLLAIVPIITAIAGVVMKMLSGOALKDKKEL 893

QY 893 EGAGKIATEAENFTVTVSLTREQKFEYMAQSLQVPRNSLRKAHIFGVFSFSTQAMMY 952

Db 894 EGAGKIATEAENFTVTVSLTREQKFEHMYDQSLQVPRNSLRKAHIFGITFSFTQAMMY 953

QY 953 FSYAGCFRFGAYLVANEFMNFQDVLVLSAIVFGAMAVGVSSFPADYAKAKYSAAHVIM 1012
Db 954 FSYAGCFRFGAYLVANESLMSPEDVLLVFSVAVFGAMAVGVSSFPADYAKAKYSAAHVIM 1013
QY 1013 IIEKSPILDSYSPHGLKPNTELEGNVTNEVFNYPTRPDIPVLOGLSLEVKKQOTLALVG 1072
Db 1014 IIEKSPILDSYSTGLKPNTELEGNVTNEVFNYPTRLDIPVLOGLSLEVKKQOTLALVG 1073
QY 1073 SSGCGKSTVVOLLERFYDPLAGSVLDGKEIKHLNVQWLAHRAHLGIYSQEPILFDCSIAEN 1132
Db 1074 SSGCGKSTVVOLLERFYDPLAGSVLDGKEIKHLNVQWLAHRAHLGIYSQEPILFDCSISEN 1133
QY 1133 IAYGDSNRVSHHEIMQAQKANIHHFIELPEKYNTIRVGDGKQTLGGQKQRIARAL 1192
Db 1134 IAYGDSNRVSHHEIMQAQKANIHHFIELPEKYNTIRVGDGKQTLGGQKQRIARAL 1193
QY 1193 VRQPHILLDEATSAIDTESKVVQVQALDKAREGRCTIVIAHRLSTIQNALDIVFQNGK 1252
Db 1194 VRQPHILLDEATSAIDTESKVVQVQALDKAREGRCTIVIAHRLSTIQNALDIVFQNGR 1253
QY 1253 VKEHGTQQLLAQKGIYFSMVSVQAGAKR 1281
Db 1254 VKEHGTQQLLAQKGIYFSMVSVQAGAKR 1282

RESULT 12
AA58187
ID AA58187 standard; Protein; 1280 AA.
XX
AC AA58187;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human G185V mutant multidrug resistance-1 (MDR-1) protein.
XX
KW Multidrug resistance; MDR-1; P-glycoprotein;
KW transmembrane efflux pump; haematopoietic stem cell; transduction;
KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW gene therapy; gene replacement; genetic defect; thalassaemia;
KW Gaucher's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW cytokine; mutan.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 185
FT /note= "This residue is Gly in the wild-type MDR-1
/note= (AA58186)"
XX
XX W09961589-A2.
XX
PD 02-DEC-1999.
XX
XX 27-MAY-1999; 99WO-US11825.
XX
XX 28-MAY-1998; 98US-0086988.
XX
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Sorrentino B, Bunting K;
XX
XX WPI; 2000-072615/06.
DR N-PSDB; AAZ49333.
XX
PT Ex vivo expansion of hematopoietic stem cells transduced with a
PT sequence encoding human multidrug resistance-1, used for bone marrow
PT transplantation -
XX
XX Example 1; Page 82-90; 113pp; English.
PS
XX This sequence represents human G185V mutant multidrug resistance
CC protein MDR-1, where the Gly residue at position 185
CC

CC of the wild-type protein (AA58186) is replaced by Val. MDR-1 is a
CC transmembrane efflux pump, responsible for the export of drugs from
CC cells, particularly cancer cells. The wild-type MDR-1 shows increased
CC resistance to etoposide and decreased resistance to vinca alkaloids
CC compared to the G185V mutant. The invention relates to transducing
CC haematopoietic stem cells with nucleic acid encoding an MDR protein
CC and culturing the modified cells. The modified haematopoietic stem
CC cells are useful in bone marrow transplantation (to reconstitute
CC haematopoietic systems in patients who have undergone chemotherapy or
CC radiation therapy) and in ex vivo gene therapy of genetic defects in
CC cells derived from haematopoietic stem cells, e.g. thalassaemia,
CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
CC cells can also be used to identify factors involved in regulating
CC proliferation and differentiation in haematopoietic stem cells.
CC Haematopoietic stem cells that express MDR-1 will be protected against
CC chemotherapeutic agents, so can be engrafted while the patient is
CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC provides sufficient cells to permit standard biochemical analysis.
CC Overexpression of MDR-1 allows cytokine-driven expansion of
CC haematopoietic stem cells by at least 10-fold compared with a maximum
CC of 4-fold in known procedures.
XX
XX Sequence 1280 AA;
QY
Query Match 90.5%; Score 5857; DB 21; Length 1280;
Best Local Similarity 90.7%; Pred. No. 0;
Matches 1163; Conservative 55; Mismatches 60; Indels 4; Gaps 4;
QY 1 MDPEGGRKGS-A-EKNFWMKGS-KKNEKKKPTVSTFAMFRYSNWLDRYMLVGTMAAI 59
Db 1 MDLEGRNGGAKKKNFKNLNNSEK-DKKEKPTVSFMSFRYSNWLDRYMLVGTMAAI 59
QY 60 IHGAALPLMLVFGNMTDSFANAGISRNKTPFVIINESITNTNQHFINHLEEMTYAY 119
Db 60 IHGAGLPLMLVFGEMTDIFANAG-NLEDIMSNITNRSNDINDTGFMMN-LEEDMTRYAY 117
QY 120 YSGIGAGLVAAAIQVSWCLAAAGQILKIRQFFHATMRQIGFVDVHDELNTRLD 179
Db 118 YSGIGAGLVAAAIQVSWCLAAAGQILKIRQFFHATMRQIGFVDVHDELNTRLD 177
QY 180 DVSKINEGIDKIGMFFQSIATFTFTGIVFTRGNKLTILVLAISPVLGSLAAWAKILS 239
Db 178 DVSKINEVIGDKIGMFFQSMATFTFTGIVFTRGNKLTILVLAISPVLGSLAAWAKILS 237
QY 240 SFTDKELLAYAKAGAAVEEVLAIRTVAFGQKKELERYKNLEAKGIGIKAITANI 299
Db 238 SFTDKELLAYAKAGAAVEEVLAIRTVAFGQKKELERYKNLEAKGIGIKAITANI 297
QY 300 SIGAAFLLIYASYALAFWYGTSLVLSSEYSIGOVLTFFSVLIGAFSIGQASPSIEAFAN 359
Db 298 SIGAAFLLIYASYALAFWYGTSLVLSSEYSIGOVLTFFSVLIGAFSIGQASPSIEAFAN 357
QY 360 ARGAAEYIFKIIDNKPSIDSYSGHGHKPDNTKGNLEFNHVSYPKREVKILGLNLKV 419
Db 358 ARGAAEYIFKIIDNKPSIDSYSGHGHKPDNTKGNLEFNHVSYPKREVKILGLNLKV 417
QY 420 QSGQTVVALVGNCGCKSTTVOLMQLYDPTDGMVCDIGDITINVRHLREITGVVQEP 479
Db 418 QSGQTVVALVGNCGCKSTTVOLMQLYDPTDGMVCDIGDITINVRHLREITGVVQEP 477
QY 480 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFTLVGERGAQLSGGQK 539
Db 478 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFTLVGERGAQLSGGQK 537
QY 540 RTAIRALVRNPKILLDEATSAIDTESAVVQVQALDKARKGRTTIVIAHRLSTVRNADV 599
Db 538 RTAIRALVRNPKILLDEATSAIDTESAVVQVQALDKARKGRTTIVIAHRLSTVRNADV 597
QY 600 TAGFDDGVIVEKGNHDELMKEGIYFKLVMTMOTRGNEIELENATGESKSDALESPDK 659
Db 598 TAGFDDGVIVEKGNHDELMKEGIYFKLVMTMOTRGNEIELENATGESKSDALESPDK 657
QY 660 SGSSLIKRRSTRSIHAPQGDRLKGTCKEDLNENPVPSFWRLKLNSTEWPFVVGIFC 719

CC of the wild-type protein (AA58186) is replaced by Val. MDR-1 is a
CC transmembrane efflux pump, responsible for the export of drugs from
CC cells, particularly cancer cells. The wild-type MDR-1 shows increased
CC resistance to etoposide and decreased resistance to vinca alkaloids
CC compared to the G185V mutant. The invention relates to transducing
CC haematopoietic stem cells with nucleic acid encoding an MDR protein
CC and culturing the modified cells. The modified haematopoietic stem
CC cells are useful in bone marrow transplantation (to reconstitute
CC haematopoietic systems in patients who have undergone chemotherapy or
CC radiation therapy) and in ex vivo gene therapy of genetic defects in
CC cells derived from haematopoietic stem cells, e.g. thalassaemia,
CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
CC cells can also be used to identify factors involved in regulating
CC proliferation and differentiation in haematopoietic stem cells.
CC Haematopoietic stem cells that express MDR-1 will be protected against
CC chemotherapeutic agents, so can be engrafted while the patient is
CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC provides sufficient cells to permit standard biochemical analysis.
CC Overexpression of MDR-1 allows cytokine-driven expansion of
CC haematopoietic stem cells by at least 10-fold compared with a maximum
CC of 4-fold in known procedures.
XX
XX Sequence 1280 AA;

Db 658 SRSSLLRKSTRSRVSGSQADKRLSTKEALDESIPPVSFWRMKLNLTUWMPYFVVGVC 717
Qy 720 AINGGLOPAFSLIESRIIGITREDPDKRONSMFSLVLGLIISFIFFLQGF 779
Db 718 AINGGLOPAFALIFSKIIGVTRIDDPETKRONSLFLFLALGLISFIFFLQGF 777
Qy 780 GKAGEITLRLRYMVRSMRLQDVSWFDDPKNTGALTRLANDAAQVKGAIGSLAVIT 839
Db 778 GKAGEITLRLRYMVRSMRLQDVSWFDDPKNTGALTRLANDAAQVKGAIGSLAVIT 837
Qy 840 QNTANLGTGIIISLIYGVWQLTLLLLAIVPIIAAGVEMKMSGOALKDKKELEGAKTA 899
Db 838 QNTANLGTGIIISFIYGVWQLTLLLLAIVPIIAAGVEMKMSGOALKDKKELEGAKTA 897
Qy 900 TEAENFRVTVSLTROKPEYMAQSLQVPIYRNSLRKAHIFGVSESIQAMMYFYAGCF 959
Db 898 TEAENFRVTVSLTROKPEYMAQSLQVPIYRNSLRKAHIFGVSESIQAMMYFYAGCF 957
Qy 960 RFGAYLVANEFMPODVLVFSIAVFGAMAVGVSSFAFDYAKAKVSAAHVIMIEKSP 1019
Db 958 RFGAYLVANEFMPODVLVFSIAVFGAMAVGVSSFAFDYAKAKVSAAHVIMIEKSP 1017
Qy 1020 IDSYPHCLKPNTLEGNTFNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSSGCGKS 1079
Db 1018 IDSYSTGLMPNTLEGNTFNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSSGCGKS 1077
Qy 1080 TVVOLLERFDPLAGSLVDIGKEIKHLNVQWLAHILGIVSQEPIILDCSIAENIAYGDN 1139
Db 1078 TVVOLLERFDPLAGSLVDIGKEIKHLNVQWLAHILGIVSQEPIILDCSIAENIAYGDN 1137
Qy 1140 RVVSHEETMOAKENIHHFIELPEKYNTRVGDKGTOLSGGQKQRIATARALVROPHIL 1199
Db 1138 RVVSHEETMOAKENIHHFIELPEKYNTRVGDKGTOLSGGQKQRIATARALVROPHIL 1197
Qy 1200 LDEATASALDTESEKVVQVQALDKAREGRTCVIAHRLSTIQNADLIIVFQNGVKVKEGHT 1259
Db 1198 LDEATASALDTESEKVVQVQALDKAREGRTCVIAHRLSTIQNADLIIVFQNGVKVKEGHT 1257
Qy 1260 QOLLAQKGIYFMSVQAGAKR 1281
Db 1258 QOLLAQKGIYFMSVQAGTKR 1279

RESULT 13
ID AAB81066 standard; Protein; 1280 AA.
AC AAB81066;
XX 25-JUN-2001 (first entry)
XX Human P-glycoprotein SEQ ID 5.
DE Cynomolgus monkey; P-glycoprotein; PGP; multidrug transporter; MDRI;
KW efflux pump; human.
XX Homo sapiens.
OS Homo sapiens.
XX WO200123565-A1.
PN 05-APR-2001.
PD 28-SEP-2000; 2000WO-US26592.
PF 28-SEP-1999; 99US-0156921.
XX 12-OCT-1999; 99US-0156818.
PR (GENT-) GENTEST CORP.
XX Stocker PJ, Steimel-Crespi DR, Crespi CL;
PI WPI; 2001-316136/33.
DR

XX Novel isolated nucleic acid encoding cynomolgus monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell -
XX
PS Claim 6; Page 68-71; 84pp; English.
XX This invention relates to a polynucleotide sequence encoding a
CC cynomolgus monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDRI is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomolgus monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents a human
CC P-glycoprotein.
XX Sequence 1280 AA;
SQ
Query Match 90.5%; Score 5857; DB 22; Length 1280;
Best Local Similarity 90.7%; Pred. No. 0;
Matches 1163; Conservative 55; Mismatches 60; Indels 4; Gaps 4;
Qy 1 MDPEGGRKGS-A-EKNFWMKMKKKKKKKKPTVSTFAMFRYSNWLDRMLYMLGTMAAI 59
Db 1 MDLEGRNGGAKKKNFFLNKSEK-DKKEKPTVSFMSFRYSNWLDRMLYMLGTMAAI 59
Qy 60 IHGAALPLMLVFGNMTDSFANAGISRNKTPPVVINESITNTQFHINLSEETTYAY 119
Db 60 IHGAGLPLMLVFGMTDIFANAG-NLEDLMSNITNRSINDTGFNMN-LEEDMTYAY 117
Qy 120 YSGIAGVLVAAYIQVSWCLAAAGQILKIRKOFFHATMRQIEGDFVDHVGELNTRLD 179
Db 118 YSGIAGVLVAAYIQVSWCLAAAGQILKIRKOFFHATMRQIEGDFVDHVGELNTRLD 177
Qy 180 DVSKINEGIDKIGMFFQSIATFFTGFIYVGTGRGKLTIVILAIISPLVGLSAAIWAKILS 239
Db 178 DVSKINEGIDKIGMFFQSIATFFTGFIYVGTGRGKLTIVILAIISPLVGLSAAIWAKILS 237
Qy 240 SFTDKELLAYAKAGAAEVLAAIRTVIAFGGKKELEKYNKLEAKGIGIKKAITANI 299
Db 238 SFTDKELLAYAKAGAAEVLAAIRTVIAFGGKKELEKYNKLEAKGIGIKKAITANI 297
Qy 300 SIGAAFLIYASALAFWYGTSLVLSSEYSIGQVLTVPFVSVLIGAFSGQASPSIEAFAN 359
Db 298 SIGAAFLIYASALAFWYGTSLVLSSEYSIGQVLTVPFVSVLIGAFSGQASPSIEAFAN 357
Qy 360 ARGAAEYEIFKIDNKPSIDSYSKSGHKPDNIKNLEKKNVHESYPSRKEVKILKGLNLKV 419
Db 358 ARGAAEYEIFKIDNKPSIDSYSKSGHKPDNIKNLEKKNVHESYPSRKEVKILKGLNLKV 417
Qy 420 QSGQTVLVGNSCGKSTTVQLMORLYDPTDGMVCIDGQDITINVRHLREITGVVSOBP 479
Db 418 QSGQTVLVGNSCGKSTTVQLMORLYDPTDGMVCIDGQDITINVRHLREITGVVSOBP 477
Qy 480 VLFATTIAENIRYGRNVTMDIEKAVKANAYDFIMKLPNKFDITLVGERGALSGGQK 539
Db 478 VLFATTIAENIRYGRNVTMDIEKAVKANAYDFIMKLPNKFDITLVGERGALSGGQK 537
Qy 540 RIAIARALVNPKILLDEATSAIDTSEAVVOVALDKARKGRTTIVIAHRLSTVRNADY 599
Db 538 RIAIARALVNPKILLDEATSAIDTSEAVVOVALDKARKGRTTIVIAHRLSTVRNADY 597
Qy 600 IAGFDDGVIVKGNHDELMKEKGIYFKLVITMOTRNEIELEENATGESKESDALEMSPKD 659
Db 598 IAGFDDGVIVKGNHDELMKEKGIYFKLVITMOTRNEIELEENATGESKESDALEMSPKD 657
Qy 660 SGSSLIKRRSTRSRSHAPQCDKRLKEDLNENVPVSWFRLKLNSTNTPWYFVVGIFC 719
Db 660 SGSSLIKRRSTRSRSHAPQCDKRLKEDLNENVPVSWFRLKLNSTNTPWYFVVGIFC 719

Db 658 SRSLIRKRRSRVSGAQRKLSTKEALDESIPVFWIRIMKLNLTWPEYVWGVC 717
 QY 720 AINGLOPAPSLIFSRIGFTFDEDETRKRONSFVLFVLGIISFIFFLQGTFF 779
 Db 718 ALINGLOPAPSLIFSRIGFTFDEDETRKRONSFVLFVLGIISFIFFLQGTFF 777
 QY 780 KGAGELTKRLRYMVFRRSMLRQDVSFWFDDPKNTTCALTRLANDAAQVKGAGISRLAVIT 839
 Db 778 KGAGELTKRLRYMVFRRSMLRQDVSFWFDDPKNTTCALTRLANDAAQVKGAGISRLAVIT 837
 QY 840 QNIANLGTGIIISLIYQWQLLILLAIIPITIAIAGVEMKMLSGALKDKKELGAGKIA 899
 Db 838 QNIANLGTGIIISLIYQWQLLILLAIIPITIAIAGVEMKMLSGALKDKKELGAGKIA 897
 QY 900 TEATENFTVSLTREQFEFYMAQSLQVYPYRNSLRKAHIFGVFSFTQAMWYSYAGCF 959
 Db 898 TEATENFTVSLTREQFEFYMAQSLQVYPYRNSLRKAHIFGVFSFTQAMWYSYAGCF 957
 QY 960 RFGAYLVANEPNFQDVLVFSIAVFGAMAVGOVSSFPADYAKAKVSAAHVIMIEKSPL 1019
 Db 958 RFGAYLVANEPNFQDVLVFSIAVFGAMAVGOVSSFPADYAKAKVSAAHVIMIEKSPL 1017
 QY 1020 IDSYSPHGLKNTLEGNTFNEVFNTPTRDIPVLOGLSLEVKKGTALVGSAGGKS 1079
 Db 1018 IDSYSPHGLKNTLEGNTFNEVFNTPTRDIPVLOGLSLEVKKGTALVGSAGGKS 1077
 QY 1080 TVOLLERFYDPLAGSLVLDGKEIKHNVOLRAHLGIVSOEPIFLDCSTAEINAYGDNS 1139
 Db 1078 TVOLLERFYDPLAGSLVLDGKEIKHNVOLRAHLGIVSOEPIFLDCSTAEINAYGDNS 1137
 QY 1140 RVVSHEETMOAAKEANIHFTETLPEKYNTRVGDGKTLQSGGQKORIAIARALVRQPHIL 1199
 Db 1138 RVVSHEETMOAAKEANIHFTETLPEKYNTRVGDGKTLQSGGQKORIAIARALVRQPHIL 1197
 QY 1200 LLDATSALDTESEKVVQVQALDKAREGTCIVIAHRLSTIONADLIYVFONGKVKHEGTH 1259
 Db 1198 LLDATSALDTESEKVVQVQALDKAREGTCIVIAHRLSTIONADLIYVFONGKVKHEGTH 1257
 QY 1260 QOLLAQKGIYFSMVSVQAGAKR 1281
 Db 1258 QOLLAQKGIYFSMVSVQAGTKR 1279
 RESULT 14
 AAE00306
 ID AAE00306 standard; Protein; 1280 AA.
 AC AAE00306;
 XX
 DT 13-JUN-2001 (first entry)
 XX
 DE Human P-glycoprotein (PGP) #1.
 XX
 KW Human; P-glycoprotein; PGP; multidrug transporter; MDR1;
 KW drug bioavailability; transgenic animal; genetic model.
 XX
 OS Homo sapiens.
 XX
 PN WO200123540-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26767.
 XX
 PR 28-SEP-1999; 99US-0156510.
 XX
 PA (GENT-) GENTEST CORP.
 XX
 PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 XX
 DR WPI; 2001-235373/24.
 XX
 PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful

PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors -
 XX
 PS Claim 16; Page 78-80; 11pp; English.
 XX
 CC The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is human P-glycoprotein (PGP). The
 CC PGP enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.
 XX
 SQ Sequence 1280 AA;
 Query Match 90.5%; Score 5857; DB 22; Length 1280;
 Best Local Similarity 90.7%; Pred. No. 0;
 Matches 1163; Conservative 55; Mismatches 60; Indels 4; Gaps 4;
 QY 1 MDPEGGKKGSA-EKNFWKMGKSKKNEKKEKPTVTFAMFRYSNWLDRYMLVGTMAAI 59
 Db 1 MDLEGRNGAKKNEFKLNKSEK-DKEKEKPTVSVFSKFRYSNWLDRYMLVGTMAAI 59
 QY 60 IHGAALPLMLVPGNMTDSFANAGISRNKTFPVIIINETSNTQHPINHEEMTYAY 119
 Db 60 IHGAGLPLMLVPGEMTDIFANAG-NLEDLSNITNRSINDTGFPMN-LEEDMTYAY 117
 QY 120 YSGIGAGLVAAVYQVSWFCLAGROILKIRKOFFHAIMROEIGWFDVHDVGLNLTLD 179
 Db 118 YSGIGAGLVAAVYQVSWFCLAGROILKIRKOFFHAIMROEIGWFDVHDVGLNLTLD 177
 QY 180 DVSKINEGDKIGMFFQSIATPFTFTGIVGTRGWKLTIVLAISVPLGSLAAWAKILS 239
 Db 178 DVSKINEVIGDKIGMFFQSIATPFTFTGIVGTRGWKLTIVLAISVPLGSLAAWAKILS 237
 QY 240 SFTDKELLAYAKAGAAVEVLAIRTVIAFGGOKKELERYNKNLEPAKIGIKKAITANI 299
 Db 238 SFTDKELLAYAKAGAAVEVLAIRTVIAFGGOKKELERYNKNLEPAKIGIKKAITANI 297
 QY 300 SIGAAFLLIYASVALAFWYGTSLVLSSEYSIGVLTFFSVLIGAFSGQASPSIEAFAN 359
 Db 298 SIGAAFLLIYASVALAFWYGTSLVLSSEYSIGVLTFFSVLIGAFSGQASPSIEAFAN 357
 QY 360 ARGAAVEIFKIIDNKPSIDSYKSGHKPNKGNLEFKNVHFSYPSRKEVKILKGLNLKV 419
 Db 358 ARGAAVEIFKIIDNKPSIDSYKSGHKPNKGNLEFKNVHFSYPSRKEVKILKGLNLKV 417
 QY 420 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDITINVRHLREITGVVQBP 479
 Db 418 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDITINVRHLREITGVVQBP 477
 QY 480 VLFATTIAENIRYGRNVTMDIEKAVKANAYDFIMKLPKPFDTLVGERGQALSGGQK 539
 Db 478 VLFATTIAENIRYGRNVTMDIEKAVKANAYDFIMKLPKPFDTLVGERGQALSGGQK 537
 QY 540 RIAIARALVRNPKILLDEATSDALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADV 599
 Db 538 RIAIARALVRNPKILLDEATSDALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADV 597
 QY 600 IAGFDGVIIVERGNHDELMKEGIYFKLVMTQTRGNEIELENATGESSEDALEMSPKD 659
 Db 598 IAGFDGVIIVERGNHDELMKEGIYFKLVMTQTRGNEIELENATGESSEDALEMSPKD 657

Thu Nov 7 09:13:12 2002

Db 718 AIIINGLOPAFAIIFSKIIGVETRIDDPETKRONSLFSLFLALGIIISFITFFLQGGTF 777
QY 780 GRAGEILTFRRLRYMVFERSMLRODVSWFDDPKNTTGALTTRLANDAAQVKGAIGSRLAVIT 839
Db 778 GRAGEILTFRRLRYMVFERSMLRODVSWFDDPKNTTGALTTRLANDAAQVKGAIGSRLAVIT 837
QY 840 QNIANLGTGIIISLIYWGWLTLALLAIVPIIAIAGVVENKMLSGQALKDKKELEGAGKIA 899
Db 838 QNIANLGTGIIISLIYWGWLTLALLAIVPIIAIAGVVENKMLSGQALKDKKELEGAGKIA 897
QY 900 TEAIENFRVSVLTREKFEYMYAQSLOVPYNSLRKKAHIFGVSPFSITQAMMYFSYAGCF 959
Db 898 TEAIENFRVSVLTREKFEYMYAQSLOVPYNSLRKKAHIFGVSPFSITQAMMYFSYAGCF 957
QY 960 RFGAYLVANEFNMQDVLVVFSAIVFAGAMAVGVSVFAPDYAKAKYSAHVIMIIIEKSPL 1019
Db 958 RFGAYLVANEFNMQDVLVVFSAIVFAGAMAVGVSVFAPDYAKAKYSAHVIMIIIEKSPL 1017
QY 1020 IDSYSPHGLKPNTEGNVTFNENVFNYPTRPDIPVLOGLSLEVKKGQTLALVSSGCCGKS 1079
Db 1018 IDSYSPHGLKPNTEGNVTFNENVFNYPTRPDIPVLOGLSLEVKKGQTLALVSSGCCGKS 1077
QY 1080 TVVOLLERFVDPLAGSVLIDGKEIKHLNQWLRAHLGIVSOBPIFLDCSIAENIAYGDN 1139
Db 1078 TVVOLLERFVDPLAGSVLIDGKEIKHLNQWLRAHLGIVSOBPIFLDCSIAENIAYGDN 1137
QY 1140 RVVSHIEIQAQAEANIHHFIETLPEKYNTRVGDKGTLQSGGQKQRIATARALVROPHIL 1199
Db 1138 RVVSHIEIQAQAEANIHHFIETLPEKYNTRVGDKGTLQSGGQKQRIATARALVROPHIL 1197
QY 1200 LDEATSALDTESEKVVQAEALDKAREGRUCIVIAHRLSTIQNADLLIVVFQNGKVKHGH 1259
Db 1198 LDEATSALDTESEKVVQAEALDKAREGRUCIVIAHRLSTIQNADLLIVVFQNGKVKHGH 1257
QY 1260 QLLAQKGIYFSMVSVQAGAKR 1281
Db 1258 QLLAQKGIYFSMVSVQAGAKR 1279

Search completed: November 6, 2002, 18:39:27
Job time : 38.4943 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: November 6, 2002, 18:45:15 ; Search time 3803.81 seconds
(without alignments)
7047.385 Million cell updates/sec

Title: US-09-672-725C-23
Perfect score: 6474
Sequence: 1 MDPEGGRKGSABKFWKMGK.....LLAQKGIVFSMVSVQAGAKR 1281

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO.spool/US09672725/runat_04112002_124341_17977/app_query.fasta_1.6435
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725.@CGN_1_11409_erunat_04112002_124341_17977 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.cm.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	6474	100.0	4279	6	AX105078 Sequence
2	6469	99.9	4279	6	AX105057 Sequence
3	6468	99.9	4279	6	AX105080 Sequence
4	6461	99.8	4279	6	AX105082 Sequence
5	6460	99.8	4045	12	AF269224 Synthetic
6	6425.5	99.3	4317	4	AF045016 Canis fam
7	6425.5	99.3	4317	6	AX105059 Sequence
8	6416.5	99.1	3934	4	CFA419568 Canis fam
9	5866	90.6	3860	6	AX322787 Sequence
10	5863	90.6	4378	6	E02326 Multidrug r
11	5862	90.5	4186	6	AX108654 Sequence
12	5859.5	90.5	4195	6	AX108656 Sequence
13	5857	90.5	3860	6	AX322789 Sequence
14	5857	90.5	4646	6	AX336420 Sequence
15	5857	90.5	4646	6	AX336708 Sequence
16	5857	90.5	4646	6	I49610 Sequence 2
17	5857	90.5	4646	9	HUMMDR1
18	5857	90.5	4669	6	I08557 Sequence 3
19	5857	90.5	6505	6	AR028671 Sequence
20	5857	90.5	8630	6	AX012320 Sequence
21	5857	90.5	8630	6	AX012321 Sequence
22	5857	90.5	9318	6	AR028672 Sequence
23	5849.5	90.4	3988	6	AX024454 Sequence
24	5849.5	90.4	4192	9	AF016535 Homo sapi
25	5830	90.1	4669	6	AR091275 Sequence
26	5815.5	89.8	4264	6	AR051647 Sequence
27	5815.5	89.8	4264	6	AR051650 Sequence
28	5797	89.5	4669	6	AR055785 Sequence
29	5687.5	87.9	3858	4	OAU78609 Ovis aries
30	5667.5	87.5	4296	10	CRUPGPI
31	5656.5	87.4	4304	10	CRUPGPI165
32	5641	87.1	4788	6	AX322793 Sequence
33	5641	87.1	4924	10	MUSMDR1A
34	5638	87.1	4356	10	MUSMDR1A
35	5634	87.0	4927	10	AF257746 Rattus no
36	5612	86.7	4323	10	AF286167 Rattus no
37	5521.5	85.3	3489	4	AB029153 Felis cat
38	5395.5	83.3	3987	10	CRUPGPI185
39	5301	81.9	4189	6	AX322791 Sequence
40	5301	81.9	4298	10	MUSMDR
41	5291.5	81.7	4233	6	AR123273 Sequence
42	5282	81.6	4280	10	CRUPGPII
43	5231.5	80.8	4254	10	RATMDRM
44	4995	77.2	3682	10	MUSPGLYIA
45	4910.5	75.8	3924	6	AX024455 Sequence

ALIGNMENTS

RESULT 1
AX105078
LOCUS
DEFINITION Sequence 22 from Patent WO0123540.
ACCESSION AX105078
VERSION AX105078.1 GI:13921228
KEYWORDS dog.
SOURCE Canis familiaris
ORGANISM Canis familiaris
linear PAT 30-APR-2001
4279 bp
DNA

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis. 1 (bases 1 to 4279) Stocker,P.J., Stelmel-Crespi,D.T., Crespi,C.L., Reif,T.C. and Patten,C.J. P-glycoproteins and uses thereof Patent: WO 0123540-A 22 05-APR-2001; GENEST CORPORATION (US) FEATURES Location/Qualifiers 1..4279 /organism="Canis familiaris" /db_xref="taxon:9615" CDS 17..3862 /note="unnamed protein product" /codon_start=1 /db_xref="GI:13921229" /translation="MDPEGGRKSAEKNFMKWKSKKNEKKKKPTVSTFAMFRYSN WLDRLYMLVGTMAAIIHGAALPLMLVPGNMTDSFANAGISRNKTFPVIINESHNNT QHFINLHEEMTTAYIYSGAGVLVAAYIQVSWCLAGRQILKIRKQFHAIMRQ EIGFDVHDVDELNRLTDDSKINEGIDKIGMFFOSIATFFTFGIVGFTRGWKLTL VLAISPIVGLSAAIWAAILKLSFDTKELLAYAKAGAAVEEVLAAIRTVIARFGQKKEL ERYNKNLSEAKGIGIKKAITANISIGAAFLIYALAFWGTSLVLSBSYSIGOVL TVFESVLGAFSIGOASPIEAFANARGAAVEIRKIDNKPISIDYSKSGHKPNIKG NLEFNKVFHSPSRKEVKLLKLNKLVQSGQTVALVNSGCGKSTTVQLMRQLDPTD GNVCDGQDIRIRINVRHUREITGVVSPQPVLFATTAENIRYGRNVTMDIERAVKE ANAYDFIMKLPNKFTLVGERGAOLSGQKQRIATARALVRNPKILLDEATSLDTE SEAVVQALDAGKGRRTIVIAHRLSTVRNADVAGDDGVIVEKGHDELMEKKGTY FKLVTMQTRGNEIELENATGESKSDALEMPKDSGLIKRSTRRSIHAPOQDR KLGTDEDLENVPPVSWRILKLNSTWEPYFVGLFCALINGLQPAFIISFRIIGI FTRDEDPETKQNSNMFSVLVLGILSIFITFFLQGFTEFGKAGELIKRLRYMFRSM LRQDVSFDDPANTGALTURLANDAQQVGAISRLVITQNIANLGTILISLYSG WOLTLULLAIPIITAIGVEMKMLSGQALKKKELEGAGKATEAENERTVVSILTR EOKFYMTAQSLVPIYRNSLRKAHIFGVSFTIQAMMTFVYAGCFREGFAYLVANEFMN FQDVLVLSATVFGAMAVGOVSFPADYAKAKVAAHVIMIEKSPILDSYSPHLKRP NTLGNGVNFVNFVNPTRDIPVLQGLSLEVKKQTLALVSSGCKSTVVOQLRFP YDPLAGSVLIDGKELKHLNVQRLRAHLGIYQEPILFDPDCSIAENIAYGDNRVVSHEE IMQAKEANIHHFIETLEPKNTRVRGDKQLSGQKQRIATARALVRNPKILLDEA TSALDTESEKVEQALDKAREGTCIVIAHRLSTIQNADLVIVFQNGKVBHGTHOOL LAQKGIFSMYSVQAGAKR" BASE COUNT 1295 a 833 c 1008 g 1143 t ORIGIN Alignment Scores: Pred. No.: 0 Length: 4279 Score: 6474.00 Matches: 1281 Percent Similarity: 100.00% Conservative: 0 Best Local Similarity: 100.00% Mismatches: 0 Query Match: 100.00% Indels: 0 DB: 6 Gaps: 0 US-09-672-725C-23 (1-1281) x AX105078 (1-4279) QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetcLys 20 Db 17 ATGGATCTCTAAGGAGCGGTAAAGGAGTGCAGAGAGAACTTCGGAAATGGGCAAA 76 QY 21 LysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40 Db 77 AAAAGTAAAAAATAGAGAAGAAAGAAAGAAACCACTGTCAGCACGTTTGCATGTTT 136 QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIlelle 60 Db 137 CGCTATTCAAATTTGGCTTGATAGGTGTATATGTTGGTGGGACAAATGGCTGCCATCATC 196 QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80 Db 197 CATGAGCTGCACACTCCCTCTCATGATGCTGGTTTTTGGAAACATGACAGATAGCTTTGCA 256 QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIlelleAsnGluSerIleThrAsn 100 Db 257 AATGCAGGAATTTCAAGAAACAAAACCTTTTCCAGTTATATTAATTAAGAAATTTACGAAC 316

QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120 Db 317 AATACACAACATTTTCATCAACCATCTGGAGGAGAAATGACACGTATGCTATATTATTC 376 QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140 Db 377 AGTGGATCGGTGCTGGCGTCTGCTTACATCCAGGTTTCATCTGGTGCTG 436 QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160 Db 437 GCAGCAGGAACACAGATACTCAAAATTAGAAAAACAATTTTTCATGCTATCATCGACAG 496 QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180 Db 497 GAGATTGGCTGGTTTGACGTACGTGGGAGCTTAACACCCGGCTCACAGAGAT 556 QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200 Db 557 GTCTCCAAAATCAATGAAGGAATTTGGCGCAAAAATTTGGAATGTTCTTTCAATCAATAGCA 616 QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220 Db 617 ACATTTTTCACCGCTTTTATAGTGGGTTTACACGTGGTGGAGCTAACCCTTGTGATT 676 QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240 Db 677 TTGCCCATCAGCCCTGTCTTGGACTTTTCAGCCGCCATCTGGGCAAAAGATACTATCTCA 736 QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260 Db 737 TTTTACTGATAAAGAACCTCTTGGCCCTATGCAAACTGGACGCTAGCTGAAGAAGTCTTA 796 QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280 Db 797 GCAGCAATCAGAACTGTGATTGCCCTTTGGAGGACAAAGAAAGAACTTGAAGAGTACAAC 856 QY 281 LysAsnLeuGluGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIleSer 300 Db 857 AAAAATTTAGAAAGAGTAAAGGAATTTGGGATAAAGAAAGCTATCACGGCAACATTTCT 916 QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320 Db 917 ATTGGTCCCGCTTTCTTATTGATCTATGATCATATGCTCTGGCTTCTGCTATGGGACC 976 QY 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340 Db 977 TCCCTTGGTCTCTCCAGTGAATATTTCTATTGGACAAGTACTACTGCTTCTTTCTGTGA 1036 QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360 Db 1037 TTAATTGGGGCTTTTAGTATTGGACAGGCTCCCAAGCATTTGAAGCATTTGCAAAAGCA 1096 QY 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380 Db 1097 AGAGGACAGCTTATGAAATCTTCAAGATATTGACAAATAAACAAGCATTTGACAGTAT 1156 QY 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400 Db 1157 TCGAAGAGTGGACATAAACCAGATAATATTAAAGGAAATTTGGAATTCAAAATGTTTAC 1216 QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420 Db 1217 TTCAGTTACCCCTTCTCGAAAGAAGTTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTTCA 1276 QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440 Db 1277 AGTGGGACAGAGTGGCGCTGGTTGGGAACAGTGGCTGCGGGGAAGACACACCGTGCAG 1336 QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460 Db 1337 CTGATGCAGAGGCTCTATGACCCACAGATGGCATGCTGCTGTTATGATGGACAGGACATT 1396 QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480

|||||
Db 1397 AGGACCAATAATGTAAGCATCTTCGGGAAATTTACTGGTGTGGTACGTACGAGGCCTGTG 1456
QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1457 TTGTTTCCACCACGATAGCTGAAACATTCGCTATGGCCGCGAAATGTCCACATGGAT 1516
QY 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1517 GAGATTGAGAAAGCTGTTAAGGAAGCCAATGCCTATGATTTTATCATGAAACTACCTAAT 1576
QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
Db 1577 AATTTGCACACTCTGGTGGAGAGAGAGGGCCCGACGCTGAGTGGTGACAGAAACAGAGA 1636
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAlaThr 560
Db 1637 ATCGCCATTGCTCGGCCCTGGTTGCGAACCCCAAGATTCTTCTGCTGGATGAGGCAACG 1696
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1697 TCAGCTCTGGACACTGAAGTGAAGCAGTGTTCAGGTGGCCCTGAGTAAGGCCAGAAAA 1756
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1757 GGCCGGACTACCATTTGTATGACTCATCGTTGTCTACACTGCTAATGCCGATGTCAAT 1816
QY 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1817 GCTGGTTTGTATGATGAGTCAATTTGTGAGAAAGAAATCATGTAACATCATGAAAGAG 1876
QY 621 LysGlyIleThrPheLysLeuValThrMetClnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1877 AAGGGCATTTACTTCAACCTTGTCCCAATGCAGACAAGAAATGAAATGTAGTTAGTAA 1936
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1937 AATGCCACTGGTGAATCCAAAAGTGAAGTGATGCCCTTGGAAATGCTCCAAAAGATTCA 1996
QY 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db 1997 GGGTCCAGTTTAATAAAGAAAGATCAACTCGCAGGAGTATACATGCACCACCAAGGCCAA 2056
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPheThr 700
Db 2057 GACAGAAAGCTTGTGTACAAAGAGAGCTTGAATGAGAAATGTACCTCCAGTTTCTCTCGG 2116
QY 701 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db 2117 AGGATTCTGAAGCTGAATCAACTGAATGGCTTATTTTGTGGTTGGTATATTTTGTGCT 2176
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2177 ATTATAACGGAGGCCCTGCAACACGACATTTTCAATAATATTTTCAAGGATTTATAGGGATC 2236
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2237 TTTTACCCGAGATGAGATCTGAAACAAACGACAGAAATAGTAAATGTTTCTGTATTG 2296
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2297 TTTCTAGTCCTTGGATTAATTTCTTTTATTAATTTTCTCCAGGGCTTCACATTTTGGC 2356
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAACTGGGGAGATCCCTCACTTAACGGCTTCGATACATGTTTTCAGATCCATGCTCAGA 2416
QY 801 GlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGGATGTCAGCTGTTGTATGACCCCTAAAACACCCTGGAGCATTTGACACACAGGCTT 2476
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
|||||

Db 2477 GCCAATGATGCGGCTCAAGCTTAAAGGGGCTATAGGTTCCAGGCTTGTCTATTACCACG 2536
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
Db 2537 AATATAGCAAAATCTTGGGACAGCATATTATATCTTAATCTATGGTTGGCAATAACA 2596
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTACTCTTAGCAATGTACCATCATTCGAATAGCAGGAGTTGTTGAATGAAATG 2656
QY 881 LeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
Db 2657 TTGCTCGGACAAAGCACTGAAGATAAGAAAGAGCTAGAAAGGAGCTGGGAAGATTGCTACA 2716
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2717 GAAGCATCGAAACCTTCGAACTGTGTCTTCTTGACTCGGAGCAGAAAGTTTGAATAC 2776
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2777 ATGTATGCACAGATTGCAAGTACCATACAGAACTCTTTGAGGAAAGCACACATCTTC 2836
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2837 GGGGTCTCATTTTATCACCCAGGCAATGATGATTTTCTATGCTGGCTGTTCGCG 2896
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
Db 2897 TTTGGTGCCTACTTGTGGCAATGAGTTCATGAACCTTCAGAGATGTTCTTTTGTGATTC 2956
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
Db 2957 TCAGCTATTGTTCTTGGTCCCATGGCAGTGGGCGAGGTCAGTTTCATTGCTCCTGACTAT 3016
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
Db 3017 GCCAAAGCCAAAGTATCAGCAGCCACGCTCATCATGATCATTTGAAAAAGCCCTCTGATT 3076
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3077 GACAGCTACAGCCCTCAGCGCCTCAAGCCCAATACGTTGGAAGGAATGTGACATTTAAT 3136
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3137 GAGGTGCTGTTCAACTATCCCACTGCACAGACATCCCCGTGCTCCAGGGGCTGAGCCTC 3196
QY 1061 GluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
Db 3197 GAGGTGAAGAAAGGGCCAGACGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGACACA 3256
QY 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
Db 3257 GTTGTTCAGCTCTACAGGCGCTTCTATGACCCCTTGGCTGGTTCAGTGTAAATGATGGC 3316
QY 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120
Db 3317 AAAGAGATAAAGCACCTGAATGTCCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAG 3376
QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
Db 3377 GAGCCCATCTCTGTGTTGACTGCACATTCGCGAAGCAATTCGCTATGGAGACACAGCCGG 3436
QY 1141 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1160
Db 3437 GTCGTATCATGAGAGATTTATGCAGGCGAGCCAGAGGCCCAACATACACCTTCATC 3496
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
Db 3497 GAGACACTCCCTGAGAAATACAACACAGTAGGAGACAAAAGGAACCCAGCTCTCTGCT 3556
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
Db 3557 GGCCAGAAACAGCGCATTCGCATAGCTCGCTCTTGTGTAGACAGCCCTCATATTTTGCCTT 3616

QY 1201 LeuAspGluAlaPheSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
|||||
Db 3617 TTGGATGAAGCTACATCAGCTTGGATACAGAAAGTGAAGAGGTGTCCAGAAAGCCCTG 3676
|||||
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
|||||
Db 3677 GACAAAGCCAGAGAGCCGACCTGCATGTGTATGCGCCACCGCTGTCCACCATCCAG 3736
|||||
QY 1241 AsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
|||||
Db 3737 AATGCAGATTAACTAGTGTGTTTCAGAAATGCAAGTCAAGAGCAGTGCACACATCAA 3796
|||||
QY 1261 GlnLeuLeuAlaGlnLysGlyIleThrPheSerMetValSerValGlnAlaGlyAlaLys 1280
|||||
Db 3797 CAGCTGCTGCCAGAAAGCATCATTTTCCATGGTCAGTGTCCAGGCTGGAGCAAAG 3856
|||||
QY 1281 Arg 1281
|||
Db 3857 CGC 3859
|||
RESULT 2
AX105057
LOCUS
DEFINITION
Sequence 1 from Patent WO0123540.
ACCESSION
AX105057
VERSION
AX105057.1 GI:13921209
KEYWORDS
dog.
SOURCE
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
1 (bases 1 to 4279)
AUTHORS
Stocker,P.J., Steimel-Crespi,D.T., Crespi,C.L., Reif,T.C. and
Patten,C.J.
P-glycoproteins and uses thereof
TITLE
Patent: WO 0123540-A 1 05-APR-2001;
JOURNAL
GENEST CORPORATION (US)
FEATURES
Location/Qualifiers
1..4279
/organism="Canis familiaris"
/db_xref="taxon:9615"
17..3862
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37718.1"
/translation="GI:13921210"
/db_xref="GI:13921210"
WLDRLYLVLVGTMAALIHGAALPLMLVFGNMTDSFANAGSRNKTFFVILINESITNNT
QHFNLHLEEMTTVAYYSGIGAGLVAAVYIQVSFWCLAAGRQILKIRKQFFHAIMRQ
EIGWFDVHVGELNRLTDDVSKINIGDKIGMFFHSIATFTFTGTVTRGWMKLT
VILATSPVLGSLAAIWAAILSSPTDKELLAYAKAAVEEVLAAIRTVIAFGGOKEL
ERYKNLEAKGIGIKKAITANI SIGAAFLLIYAYALATWYGTSLVLSSEYSTGOVL
TVFSPVLGASIQGASPSIEAFANARCAAYEIPKIDNKPSSIDSYSKSHKPDNKG
NLEKFNHFSIPSRKEVKILKLVQSGOTVALVNSGCGSTTVQLMQRLLDPD
GMVCDIGDQIRTIINVRHLREITGVVSOEPLVFATTAENIRYGRENTVMDIEKAYKE
ANAYDFIMKLPKEDTILVGERAQLSGQKQRIARALVRNPKILLDDEATSLDTE
SEAVQVAKQSLQVYPYRNSLRKAHIFGFSIQAAMFYFAGCFRFGALVANEPMN
FODVLYFAQLVFCAMAVGVSSFPADYAKAVSAAHVIMIIEKSPILDSYSPHGLK
KLCNKEDLNENVPVSWRILKLNSTENPYFVVGIFCAILINGLOPAPFISFRIGI
FTREDPETRKQNSNMISVLFVLGIISIFTFQFGTGGKAGELLKRLKRYWVFRSM
LRQVSDPDDPKNTGALTTLANDAAQVKAIGSRNAVITQNIANGTIIISLYG
WOLRLLLAIVPIITAIGVVMKMLSGALKDKKEGAKIATEENRPTVVSLLTR
RQKEFYMAQSLQVYPYRNSLRKAHIFGFSIQAAMFYFAGCFRFGALVANEPMN
FODVLYFAQLVFCAMAVGVSSFPADYAKAVSAAHVIMIIEKSPILDSYSPHGLK
NPLEGNVTENVEVNYTRPDI PVYLGISLEVKQQLALVSGSGCKSKTVQVLLERF
YDPLAGSVLIDGKEIKHLNWLRAHIGVISOEPILFDCSIAENIATGDNRSVSHDEE
IMQAANKHHHFTETIPEKYNTRVGDGTQLSGQKQRIARALVRNPKILLDDEA
TSALDTESEKVVQALDKREGRTICVIAHRLSTIQNLIDLVVFNQGVKVEHGTQOOL
LAQGIYFSVSVQAGAKR"
BASE COUNT 1294 a 834 c 1008 g 1143 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6469.00 Matches: 1280
Percent Similarity: 99.92% Conservative: 0
Best Local Similarity: 99.92% Mismatches: 1
Query Match: 99.92% Indels: 0
DB: 6 Gaps: 0
US-09-672-725c-23 (1-1281) x AX105057 (1-4279)
QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
|||||
Db 17 ATGGATCCTGAAGAGAGCCGTAAGGGAGTGCAGAGAAAGAACTCTCTGGAATAATGGGCAA 76
|||||
QY 21 LysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
|||||
Db 77 AAAAGTAAAAAATGAGAGAAAGAAAGAAACCACTGTCAGCAGCTTTGCAATGTTT 136
|||||
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 60
|||||
Db 137 CGCTATTCAAATTTGGCTTGATAGTTGTATATGTTGGTGGGACAATGGCTGCATCATC 196
|||||
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
|||||
Db 197 CATGGAGCTGCACCTCCCTCTCATGATGCTGCTTTTGGAAACATGACAGATAGCTTTGCA 256
|||||
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
|||||
Db 257 AATGCAGAAATTTCAAGAAACAAAACCTTTCCAGTTTATAATTAATGAAGATATTACGAAC 316
|||||
QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120
|||||
Db 317 AATACACAACATTTTCATCAACCATCTCGAGGAGAAATGACCAGTATGCTTATTATTAC 376
|||||
QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
|||||
Db 377 AGTGGATCGCTGCTGGCTGGCTGGCTGTACATCCAGGTTCAITCTGGTGGCTG 436
|||||
QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
|||||
Db 437 GCAGCAGAAAGACAGATACTCAAAATTTAGAAAACAATTTTTCATGCTATCATGCGACAG 496
|||||
QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
|||||
Db 497 GAGATTGGCTGGTTTGACGTGACGTGGGAGCTTAACCCCGGCTCACAGACGAT 556
|||||
QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIleAla 200
|||||
Db 557 GTCTCCAAAATCAATGAAGAAATGGCGCAAAATTTGGAATGTTCTTCACTCAATAGCA 616
|||||
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
|||||
Db 617 ACATTTTTCACCGGTTTATAGTGGGGTTTACAGTGGTGGTGAAGCTAACCCCTGTGATT 676
|||||
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
|||||
Db 677 TTGGCCATCAGCCCTGTTCTTGGACTTTTCAGCGCCCATCTGGGCAAGATACTATCTTCA 736
|||||
QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyValAlaGluGluValLeu 260
|||||
Db 737 TTTACTGATAAAGAACTCTTGGGCTATGCAAAAAGCTGGAGCAGTAGCTGAAGAAGCTTA 796
|||||
QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyrAsn 280
|||||
Db 797 GCAGCAATCAGAACTGTGATTGCTTTGGAGGACAAAAGAAAGAACTTGAAAGGTACAAC 856
|||||
QY 281 LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
|||||
Db 857 AAAAATTTAGAAGAACTAAAGAAATTTGGGATAAAGAAAGCTATCACGGCCAACTTTCT 916
|||||
QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
|||||
Db 917 ATTGGTGGCCGCTTCTTATTGATCTATGCATCATATGCTCTGCTGCTTCTGGTATGGGACC 976
|||||

QY	321	SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal	340
DB	977	TCCTTGGTCTCTCCAGTGAATATCTATTGGACAAGTACTCACTGTCTCTCTTTCTGTA	1036
QY	341	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	360
DB	1037	TTAATTGGGGCTTTAGTATTGCACAGGCATCCCAAGCATTTGAAGCATTTGCAAAACGCA	1096
QY	361	ArgGlyAlaLatyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	380
DB	1097	AGAGGACAGCTTATGAATCTTCAAGATAATTGACAATAAACCAAGCATTTGACAGCTAT	1156
QY	381	SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	400
DB	1157	TCGAAGAGTGGACATAAACACAGATAATATTAAGGAAATTTGGAAATTCAAAATATTGCAC	1216
QY	401	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	420
DB	1217	TTTCAGTTACCTCTCTGAAAGAAGTTAAGATCTTTAAGGGTCTCAACTGAAGGTTACG	1276
QY	421	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln	440
DB	1277	AGTGGCAGACAGTGGCGCTGTTGGACACAGTGGCTGCGGGAAGAGCAGCACCGTCGAC	1336
QY	441	LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	460
DB	1337	CTGATGCAGAGGCTCTATGACCCACACAGATGCATGGTCTGTATTGATGGACAGACATT	1396
QY	461	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	480
DB	1397	AGGACCAATAATGTAGGCATCTTCGGGAATTAATCTGGTGGTGAGTCAGAGCCGTG	1456
QY	481	LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	500
DB	1457	TTGTTTCCACCACGATAGCTAAAAACATTCGCTATGCGCGCAAAATGTCACCATGGAT	1516
QY	501	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	520
DB	1517	GAGATTGAGAAAGCTGTTAAGGAAGCCAATGCTGTATGATTTATCATGAACACTACCTAAT	1576
QY	521	LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg	540
DB	1577	AAATTTACACTCTGTTGGAGAGAGAGGGCCAGCTGAGTGTGGACAGAAACACAGA	1636
QY	541	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	560
DB	1637	ATCGCCATTGCTCGGGCCCTGGTTCGCACACCCAAAGATCTCTCGCTGGATGAGCAACG	1696
QY	561	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	580
DB	1697	TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGTGGCCCTGGATTAAGGCCAGAAA	1756
QY	581	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	600
DB	1757	GGCCGGACTACCATTTGTGATAGCTCATCGTTTGTCTACAGTTCGTAATGCGGATGTCAAT	1816
QY	601	AlaClyPheAspGlyValIleValGluLysGlyLysHisAspGluLeuMetLysGlu	620
DB	1817	GCTGGTTTTGATGATGGAGTCAATTGTGGAGAAGAAATCATGATGAATCATGAAAGAG	1876
QY	621	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	640
DB	1877	AAGGCCATTTACTTCAACTTGTCAATATGCACAGACAGAGAAATGAAATTTAGTTAGTAAG	1936
QY	641	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	660
DB	1937	AATGCCACTGGTCAATCCAAAAGTGAAGTGTATGCTTTGGAATGTCTCCAAAAGATTCA	1996
QY	661	GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln	680
DB	1997	GGGTCAAGTTTATAAAAAAGAGATCAACTCGCAGGAGTATACATGCACCAAGGCCAA	2056

QY	681	AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp	700
DB	2057	GACAAAGCTTGCTACAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCTCTTCGG	2116
QY	701	ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla	720
DB	2117	AGGATTCTGAAGCTGAACCTCAACTGAATGGCCTATTATTGTGGTGGTATATTTTGGCT	2176
QY	721	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	740
DB	2177	ATTATAACGAGGCGCTGCACACAGCATTTTCAATAATATTTCAAGGATTTATAGGATC	2236
QY	741	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	760
DB	2237	TTTACCCGAGATGAGGATCCTGAACAAACACAGAGAATAGTAACATGTTTCTGTATTG	2296
QY	761	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	780
DB	2297	TTTCTAGTCCTTGGAAATATTCTTTTATATACATTTTCTCCAGGGCTTCACATTTGGC	2356
QY	781	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	800
DB	2357	AAAGCTGGGAGATCCTCACTAAGCGGCTTCGATACATGGTTTTCAGATCCATGCTGAGA	2416
QY	801	GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	820
DB	2417	CAGGATGTCAGCTGGTTGATGACCTAAAAACACCACCTGGAGCATTCACAACACGGCTT	2476
QY	821	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	840
DB	2477	GCCAATGATCGGCTCAAGTTAAAGGGGCTATAGGTTCAGGCTTGCTGTCATTTACCACG	2536
QY	841	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr	860
DB	2537	AATATAGCAATCTTGGACAGCATATATATATCTTAATCTATGGTGGCAATTAAACA	2596
QY	861	LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet	880
DB	2597	CTTTTACTCTTAGCAATGTGACCACATCATGCAATAGCAGGAGTTGTGAAATGAAATG	2656
QY	881	LeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAlaThr	900
DB	2657	TTGTCTGGACAGCAGCTGAAGATAAGAAGAGCTAGAAGAGCTGGGAAGTTGCTTACA	2716
QY	901	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr	920
DB	2717	GAAGCCATCGAAACCTCCGAACTGTGTTCTTTGACTCGGGACGAGAAGTTTGATATC	2776
QY	921	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	940
DB	2777	ATGTTATGCACAGAGTTTGC AAGTACCATTACAGAAACTCTTTGAGAAAGACACATCTTC	2836
QY	941	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	960
DB	2837	GGGTGCTCATTTCTATCACCCAGGCAATGATGATTTTCTATGCTGGCTGTTTCCGG	2896
QY	961	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe	980
DB	2897	TTTGGTGCCTACTTGGTGGCAATGAGTTCATGAACTTTCAGGATGTTCTTTGGTATTC	2956
QY	981	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	1000
DB	2957	TCAGCTATTGCTTTGGTGCATGGCAGTGGGCGAGGTCAGTTTCAATTTGCTCCTGACTAT	3016
QY	1001	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1020
DB	3017	GCCAAAGCCAAAGATATCAGCAGCCACGCTCATCATGATCATTTGAAAGAACCCCTGATT	3076
QY	1021	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1040
DB	3077	GACAGCTACAGCCCTCACGGCTCAAGCCCAATACGTTTGGGAAGGAATGTGACATTAAT	3136
QY	1041	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1060

|||||
Db 497 GAGATTGGCTGGTTTACGCTGGCATGACGTTGGGAGCTTAACACCGGCTCACAGCGAT 556
QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
Db 557 GTCTCCAAAATCAATCAAGGAATTGGCGACAAAATTGGGAATGTTCTTTTCAATCAATAGCA 616
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
Db 617 ACATTTTTCACCGGTTTATAGTGGGGTTTACACGTTGGTGGGAAGCTAACCCCTTGTAAT 676
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
Db 677 TTGGCCATCAGCCCTGTTCTTGACATTTTCAGCCGCCATCTGGGCAAGATACATACTTCA 736
QY 241 PheThrAspLysGlnLeuLeuAlaTyraLysAlaGlyAlaValAlaGluGluValLeu 260
Db 737 TTTACTGATAAAGAACTCTTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 796
QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGlnLeuGluArgTyrAsn 280
Db 797 GCAGCAATCAGACTCTGTGATGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAAC 856
QY 281 LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
Db 857 AAAAATTTAGAGAAGCTAAAGGAATTTGGGATAAAGAAAGCTATCAGGCCAACATTTCT 916
QY 301 IleGlyAlaAlaPheLeuLeuIleTyraLysTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
Db 917 ATTTGGTCCGCTTCTTATTGATCTATGCATCATATGCTCTGGCTTCTTGGTATGGGACC 976
QY 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
Db 977 TCCTTGTGCTCTCCAGTGAATATCTTATTGGACAAGTACTACTGTCCTTCTTTCTGTA 1036
QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db 1037 TTAATTTGGGCTTTAGTATTGGACAGCATCCCAAGCATTTGAAGCATTTGCAACGCCA 1096
QY 361 ArgGlyAlaIaTyraGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
Db 1097 AGAGGACAGCTTATGAAATCTTCAAGATAATTGACAATAAACAAGCATTTGACAGTAT 1156
QY 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1157 TCGAAGAGTGGACATAACAGATAATATTAAAGGAAATTTGGAATTCAAAATAGTTTAC 1216
QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1217 TTCAGTTACCTTCTCGAAGAAGTTAAGATCTTAAGGGTCTCAACCTGAAGGTTACAG 1276
QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1277 AGTGGCAGACAGTGGCGCTGTTGGGAACAGTGGCTGCGGGAAGAGACGACCGTGCAG 1336
QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db 1337 CTGATGACAGAGGCTCTATGACCCACAGATGGCATGCTGTAATGATGGACAGGACATT 1396
QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db 1397 AGGACCATAATGTAAGGCATCTTCGGGAATTTACTGGTGGTGAGTCAGGAGCCCTGTG 1456
QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1457 TTGTTTGGCCACCAGTAGCTGAAACATTCGCTATGGCCGCGAAAATGTCAACCATGGAT 1516
QY 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyraPheIleMetLysLeuProAsn 520
Db 1517 GAGATTGGAAGCTGTTTAAAGGAAGCCCAATGCCCTATGATTTTATCATGAAGAACTACCTAAT 1576
QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
|||||

Db 1577 AAATTTGACACTCTGTTGGAGAGAGAGGGCCCGAGCTGAGTGTGGACAGAAACAGAGA 1636
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
Db 1637 ATCGCCATTGTCGGGCCCTGGTTGCGCAACCCCAAGATTCTTCGTGGATCAGGCAACG 1696
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1697 TCAGCTCTGGACACTGAAGTGAAGCAGTGGTTTCAGGTGGCCCTGGATAAGGCCAGAAA 1756
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1757 GGCCGGACTACCATTTGTATAGTCTCATGTTTGTCTACAGTTCGTAATGCCGATGTCAAT 1816
QY 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1817 GCTGTTTTGTATGATGGAGTCATTTGTGGAAAGGAATCATGATGAACATCATGAAGAG 1876
QY 621 LysGlyIleTyraPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1877 AAGGCCAATTTACTTCAAACTTGTCAATGCAGACAAGAGGAATGAAATGAGTTAGAA 1936
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1937 AATGCCACTGGTGAATCCAAAAGTGAAGTGAATGCTTGGAAATGCTCCAAAAGATTCA 1996
QY 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db 1997 GGGTCAGTTTAAATAAAGAAGATCACTCCAGAGATATACATGCACCAACAGGCCAA 2056
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
Db 2057 GACAGAAAGCTGGTACAAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCCCTCTCG 2116
QY 701 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db 2117 AGGATTCTGAAGCTGAACCTCAACTGAATGGCCTTATTGTGGTTGGTATATTTTGTGCT 2176
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2177 ATTATAAACGGAGCCCTGCAACAGCAATTTTCAATAATATTTTCAAGGATATAGGATC 2236
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2237 TTTACCCGAGATGAGGATCCTGAAACAAAACACAGAAATAGTAACATGTTTCTGTATTG 2296
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2297 TTTCTAGTCTTTGGAAATATTTCTTTTATACATTTTCTCCAGGGCTTCACATTTGGC 2356
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAAGCTGGGAGATCTCTACTAAGCGCTTCGATACATGTTTTCAGATCCATGCTGAGA 2416
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGGATGTCAGCTGGTTTGTATGCCCTAAACACACCACTGGAGCATTCACACAGGCTT 2476
QY 821 AlaAsnAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2477 GCCAATGATGGCGCTCAAGTTAAAGGGGTATAGGTTCCAGGCTTGTCTCATTTACCCAG 2536
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyraGlyTrpGlnLeuThr 860
Db 2537 AATATAGCAAAATCTGGGACAGCATTAATATATCTTAATCTATGTTGGCAATTAACA 2596
QY 861 LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTCTACTCTTAGCAATTTGATCCCATCATTTGTAATAGCAGGAGTTGTTGAAATGAAAATG 2656
QY 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluAlaGlyLysIleAlaThr 900
Db 2657 TTGCTCGACAGCACTGAAAGATAAGAAAGAGCTAGAAGGAGCTGGGAAGATTGCTACA 2716

Db	2237	TTTTACCCGAGATGAGGATCTCGAACAACAAACGACAGATAATAGTAACTGTTTCTGTATTG	2296
Qy	761	PheLeuValLeuGlyIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	780
Db	2297	TTTCTAGTCTTGGAAATATTCTTTTATTACATTTTCTCCAGGCTTCACATTGGC	2356
Qy	781	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	800
Db	2357	AAAGCTGGGAGATCCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTGAGA	2416
Qy	801	GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	820
Db	2417	CAGGATGTCAGCTGGTTGTGATGACCCCTAAACACACCATGGAGCATTTGACAAACGAGCTT	2476
Qy	821	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	840
Db	2477	GCCAAATGCGGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTCTGCTCATACCCAG	2536
Qy	841	AsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThr	860
Db	2537	AATATAGCAAACTCTGGGACAGGCAATTATATATCTTATCTATGTTGGCAATTAACA	2596
Qy	861	LeuLeuLeuLeuAlaIleValProIleAlaIleAlaIleValValGluMetLysMet	880
Db	2597	CTTTTACTCTTAGCAATTGTACCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAAAAATG	2656
Qy	881	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr	900
Db	2657	TTGTCTGCACAAGCACTGAAAGATAAGAAAGACGTAGAGGAGCTGGGAAGATTGCTACA	2716
Qy	901	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr	920
Db	2717	GAAGCCATCGAAAACCTCCGAACCTGTTCTTTGACTCGGAGCAGAAGTTTGAATAC	2776
Qy	921	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	940
Db	2777	ATGTATGCACAGAGTTTGCAAAGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTTC	2836
Qy	941	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	960
Db	2837	GGGTCTCATTTTCTATCACCAGCAATGATGTAATTTTCTATGCTGCTGTTCCGG	2896
Qy	961	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe	980
Db	2897	TTTGGTGCCTTACTTGGTGCAATGAGTTTCATGAACCTTTTCAGGATGTTCTTTTGGTATTTC	2956
Qy	981	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	1000
Db	2957	TCAGCTATTGCTTTTGGTGCCATGGCAGTGGGCGAGGTGAGTTCTCTCTGACTAT	3016
Qy	1001	AlaLysAlaLysValSerAlaAlaHisValIleMetIleGluLysSerProLeuIle	1020
Db	3017	GCCAAAGCCAAATATACAGAGCCCGTCATCATGATCATTTGAAAGAGCCCTCTGATT	3076
Qy	1021	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1040
Db	3077	GACAGCTACAGCCCTCACGGCTCAAGCCAAATACGTTTGAAGGAAATGTGACATTTAAT	3136
Qy	1041	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1060
Db	3137	GAGTCTGTTTCACTATCCACTCGACACAGATCCCCGCTCCAGGGCTGAGCCTC	3196
Qy	1061	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1080
Db	3197	GAGGTGAAGAAGGCCAGAGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGAGCACA	3256
Qy	1081	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1100
Db	3257	GTGTGTCAGCTCTAGAGCGCTTCTATGACCCCTTGGCTGTTGCTAGTCTAATTCATGGC	3316
Qy	1101	LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1120
Db			
Db	3317	AAAGAGATAAAGACACCTGAATGTCAGTGGCTCCGAGCACACCTGGCATCTGTCTCAG	3376
Qy	1121	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1140
Db	3377	GAGCCATCTCTTTTACTGTCAGCATTCGCGAGAATTTGCCATATGGAGACAACGCCGG	3436
Qy	1141	ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle	1160
Db	3437	GTCGTATCACATGAAGAGATTGTGCGAGCGCAAGAGGCAACATACACCACTTCATC	3496
Qy	1161	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1180
Db	3497	GAGACACTCCCTTGAGAAATACAAACACCAAGAGTAGGAGCAAAAGGAAACCACTCTCTGGT	3556
Qy	1181	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1200
Db	3557	GGCCAGAAACAGCGCATTTGCCATAGCTCGCGCTCTTTGTAGACAGCCCTCATATTTTGCTT	3616
Qy	1201	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1220
Db	3617	TTGGATGAAGCTTACATCAGCTCTGATACAGAAAGTGAAGAGTTGTCCAAGAGCCCTG	3676
Qy	1221	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1240
Db	3677	GACAAAGCCAGAGAGCGCCACCTGCTATTGTATCGCCACCCGCTTGTCACCATCCAG	3736
Qy	1241	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1260
Db	3737	AATGCAGATTAAATAGTGGTGTTCAGAAATGGCAAGTCAAGGAGCATGGCACACATCAA	3796
Qy	1261	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys	1280
Db	3797	CAGCTGCTGGCCAGAAAAGGCATCTATTTTCCATGCTGCTCCAGGCTGGAGCAAG	3856
Qy	1281	Arg 1281	
Db	3857	CGC 3859	
RESULT	5		
LOCUS	AF269224	4045 bp DNA linear SYN 04-JUL-2000	
DEFINITION	Synthetic construct Canis familiaris his-tagged-multidrug resistance glycoprotein gene, complete cds.		
ACCESSION	AF269224		
VERSION	AF269224.1		
KEYWORDS	AF269224.1 GI:8926216		
SOURCE	synthetic construct		
ORGANISM	synthetic construct		
REFERENCE	1 (bases 1 to 4045)		
TITLE	Roulet,A.C., Puel,O., Gesta,S., Alvinerie,M. and Pineau,T.J.		
JOURNAL	Analysis of dog MDR1 p-glycoprotein		
REFERENCE	2 (bases 1 to 4045)		
AUTHORS	Roulet,A.C., Puel,O., Gesta,S., Alvinerie,M. and Pineau,T.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (19-MAY-2000) Pharmacology, INRA, 180 Chemin de Tournefeuille, BP3, Toulouse 31931, France		
FEATURES	Location/Qualifiers		
source	1..4045		
	/organism="synthetic construct"		
	/db_xref="taxon:32630"		
	/note="original mRNA isolated from Beagle dog colon"		
	101..3964		
	/codon_start=1		
	/transl_table=11		
	/product="his-tagged-multidrug resistance glycoprotein MDR1"		
	/protein_id="AAF81747.1"		
	/db_xref="GI:8926217"		
	/translation="MHHHHHDPGGKKGSAEKNFKMKKKKNEKKKPTVSTFA MFYSWNLDRLYMLVCTMAAIIHGAALPLMLVFGNWTDSFANSGSRNKTTPVINE SITNNTQHFNLHEEEMTITATYISGIGAGVLVAATYIQVSWCLAAGROLIKRKQFF		
CDS			

HAIMRQIGFVDFVHVGELNRLTRDDVSKINEGIDGKIMGFQSIATFFTFGIVGTR
GWKLTILAI5PVLGSAAIWAKILSSFTDKELLAYAKAGAAVEEVLAIIRTVIARF
GOKELERYKNLEAKRIGIKKAITANISIGAAFLIIVASYALAEWYGFSLVLSSEY
SIGOVLTVF5VLIGAF5ICQASPSIEAFANARGAAIEPKIIDNKP5IDSYSKSHK
PDNIKGPEKNH5F5PSRKEVKILGLNKLKVGSGOTVALVNSGCGKSTTVOLMOR
LYDPTDGMV5ICDQDIRTINVRHLREITGV5QEPVLFAITIAENTRYGRENVTDEI
EKAVKEANAYDFIMKLPNKEDTLVGERGAOLSGQKORIAIRALVRNPKILLDEAT
SALDTESEAVVQALDKARKGR5TVIAHRL5VRNADVIAGDFDGVIVBKGNHDELM
KEKGIYFKLVTMOTRNEIELEENATGSK5ESDALEMPKDSG5SLIKRSTR5IHA
POGDRKLTKEDL5ENVP5P5FW5RILKLN5T5WP5PV5GIF5CALING5LOP5AF5IF
SRIGIF5TRED5ETK5R5N5MF5VL5GLI5I5FT5F5Q5FT5FK5AG5IL5FK5RL5
M5FR5ML5R5D5F5D5PK5NT5GAL5T5TRL5AND5A5VK5A5G5L5AV5T5QNT5IAN5T5GII
ISL5IY5Q5L5L5I5L5I5V5I5I5A5GV5EM5K5L5G5Q5AL5DK5KE5L5E5G5K5IA5F5A5EN5FT
V5SL5TR5EQ5K5E5V5A5S5LO5VP5YR5N5SL5K5AH5IF5GV5FS5ITO5AM5Y5F5V5AG5FR5G5AY5IV
AN5EPM5NF5OD5V5L5F5SA5IV5FC5AM5AV5GV5SS5P5AD5Y5AK5AV5AA5HV5IM5IEK5PL5ID5YS
PH5GL5K5PNT5L5GN5TV5F5N5V5Y5TR5D5IP5VL5Q5L5S5L5EV5K5G5Q5L5AL5V5G5SG5C5K5SV5V
QL5L5RF5D5I5AG5S5L5ID5G5K5HL5N5V5M5L5RA5H5IG5IV5Q5EP5IL5F5D5CS5IA5EN5I5AT5Y5R5PHI
V5SH5E5T5M5Q5AK5AN5IH5IF5I5ET5PE5K5Y5TR5V5G5DK5Q5L5SG5G5K5OR5IA5I5AR5I5AL5Y5R5PHI
LL5LD5EAT5SAL5D5E5K5Y5Q5B5AL5DK5ARE5G5RT5C5IV5I5A5H5RL5ST5IQ5N5AD5L5IV5P5Q5N5K5V5KEH
GTH5Q5LL5A5Q5K5GI5V5F5M5IS5V5Q5AG5AK5R5"

BASE COUNT 1171 a 850 c 986 g 1038 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4045
Score: 6460.00 Matches: 1278
Percent Similarity: 99.92% Conservative: 1
Best Local Similarity: 99.84% Mismatches: 1
Query Match: 99.78% Indels: 0
DB: 12 Gaps: 0

US-09-672-725C-23 (1-1281) x AF269224 (1-4045)

QY 2 AspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLysLys 21
DB 122 GATCTGAAGGAGCGCTAAGGGAGTGCAGAGAAGAACTTCTCGAAATGGGCAAAA 181
QY 22 SerLysLysAsnGluLysGluLysLysProThrValSerPheAlaMetPheArg 41
DB 182 AGTAAAAAATGAGAGAAAGAAAGAACCACTGTCACACAGTTTGGCAATGTTTCGC 241
QY 42 TyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleHis 61
DB 242 TATTCAAATGGCTTGATAGTTGATATGTTGGTGGGACAAATGGCTGCCATCATCAT 301
QY 62 GlyAlaAlaLeuProLeuMetLeuValPheGlyAsnMetThrAspSerPheAlaAsn 81
DB 302 GGAGCTGCACCTCCTCATGATGCTGTTTGGAAACATGACAGATAGCTTTGCAAT 361
QY 82 AlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsn 101
DB 362 CGAGAAATTCAGAAACAAAATTTCCAGTTATAATTAATGAAGTATTACGAACAT 421
QY 102 ThrGlnHisPheIleAsnHisLeuGluGluMetThrThrThrAlaTyrTyrSer 121
DB 422 ACACAAATTTTCATCAACCATCTGGAGGAGAAATGACCAGTATGCTATTATTACAGT 481
QY 122 GlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeuAla 141
DB 482 GGGATCGGTGGCGTCTGGTGGCTTACATCCAGGTTTATTCTGGTGGCTGGCA 541
QY 142 AlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGlu 161
DB 542 GCAGGACACAGATACCTAAAATAGAAAACAATTTTTCATGCTATCATCGCAGGAG 601
QY 162 IleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspVal 181
DB 602 ATTTGGCTGGTTGACCTGACCTGACCTGGGAGCTTAACACCGGCTCACAGACGATGC 661
QY 182 SerLysIleAsnGluLysIleGlyAspLysIleGlyMetPhePheGlnSerIleAlaThr 201
DB 662 TCCAAAATCAATGAAGGAATTTGGGACAAAATTTGGAATGTTCTTCAATCAATAGCAACA 721

QY 202 PhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIleLeu 221
DB 722 TTTTTCACCGGTTTATAGTGGGTTTACACGGTGGTGGAAAGCTAACCTTTGTGATTTG 781
QY 222 AlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPhe 241
DB 782 GCCATCAGCCCTCTCTTGGACATTCAGCGCCATCTGGGCAAGATACATCTTCATT 841
QY 242 ThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeuAla 261
DB 842 ACTGATAAAGACTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGCTTAGCA 901
QY 262 AlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsnLys 281
DB 902 GCAATCAGAACTCTGATGGCTTTGGAGGACAAAGAAAGAACTTCAAAGGTACAAACA 961
QY 282 AsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSerIle 301
DB 962 AATTTAGAAGAACTAAAAGAAATTTGGATAAAGAAAGCTATCACGCCCAACATTTCTAT 1021
QY 302 GlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSer 321
DB 1022 GTGCGCCCTTCTATTGATCTATGATCATATGCTTCTGGCTTCTGTTATGGACCTCC 1081
QY 322 LeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerValLeu 341
DB 1082 TTGTCCTCTCCAGTGAATATTCTATTGGACAAAGTACTCACTGCTCTTCTTCTGTATTA 1141
QY 342 IleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArg 361
DB 1142 ATTGGGGCTTTAGTATTGACAGCATCCCAAGCATTTGAAGCATTTGCCAAACGCAAGA 1201
QY 362 GlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSer 381
DB 1202 GGACAGCTTATGAATCTTCAAGATAATTGACAAATAAACCAAGCAATTGCACGATTCG 1261
QY 382 LysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPhe 401
DB 1262 AAGAGTGCACATAAACACAGATAATATTAGGGAAATTTGGAATTTCAAAAATGTTCACTTC 1321
QY 402 SerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSer 421
DB 1322 AGTTACCTCTCTCGAAAGAAAGTTAAAGATCTTAAAGGGTCTCAACCTGAAGGTTTCAGAT 1381
QY 422 GlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeu 441
DB 1382 GGCAGACAGCTGGCGTGGTGGACACAGTGGCTGGGGAGAGACAGACCCTGGAGCTG 1441
QY 442 MetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArg 461
DB 1442 ATGCAGAGGCTTATGACCCACAGATGGCATGCTGCTGTTATGATGGACAGGACATTAG 1501
QY 462 ThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeu 481
DB 1502 ACCATAATGTAAAGGCATCTTCGGGAAATTTACTGGTGGTGGTGGTGGTGGTGGTGG 1561
QY 482 PheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAspGlu 501
DB 1562 TTTGCCACACAGATAGCTGAAACATTCGCTATGGCGCGGAAATGTCACCATGGATGAG 1621
QY 502 IleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsnLys 521
DB 1622 ATTGAGAAAGCTGTTAAGGAAGCAATGCCATGATGATGATGATGATGATGATGATGAT 1681
QY 522 PheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArgIle 541
DB 1682 TTTGACACTCTGTTGGAGAGAGAGGGCCAGCTGAGTGGTGGTGGTGGTGGTGGTGGTGG 1741
QY 542 AlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThrSer 561
DB 1742 GCCATTGCTGGGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1801
QY 562 AlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLysGly 581

Db 1802 GCTCTGGACACTGAAGAGTGAAGCAGTGGTTCAGGTGGCCCTGGATAAGGCCAGAAAAGGC 1861
Qy 582 ArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla 601
Db 1862 CGGACTACCAATGTGATAGCTCATCGCTTGTCTACAGTTCGTAATGCCGATGTCAATGCT 1921
Qy 602 GlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGluLys 621
Db 1922 GGTTTTGATGATGGAGTCATTTGGGAGAAAGGAAATCATGATGAATCATGAAGAGAG 1981
Qy 622 GlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGluAsn 641
Db 1982 GGCAATTTACTTCAAACTTGTCAAAATGCAGACAAAGAGAAATCAAAATTCAGTTAGAAAT 2041
Qy 642 AlaThrGlyLysSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSerGly 661
Db 2042 GCCACTGGTGAATCCAAAGTGAAGTGAATGATGCTTGGAAATGCTCCAAAGATTCAGGG 2101
Qy 662 SerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAsp 681
Db 2102 TCCAGTTTAATNAAGAAGATCACTCGCAGAGTATACATGCACCACCAAGCCCAAGAC 2161
Qy 682 ArgGlyLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrpArg 701
Db 2162 AGAAAGCTTGGTACAAAAGAGGACTTGAATGAGAAATGTACCTCCAGTTTCTCTCTGGAGG 2221
Qy 702 IleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAlaIle 721
Db 2222 ATTTCTGAAGCTGAACCAACTGAATGGCCCTATTTTGTGGTATATATTTGTGCTAAT 2281
Qy 722 IleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePhe 741
Db 2282 ATAAACGGAGCCTGCAACCAAGCATTTTCAATAATATTTTCAAGGATATAGGATCTTT 2341
Qy 742 ThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe 761
Db 2342 ACCCGAGATGAGATCCTGAAACAAAACGACAGAAATAGTAACATGTTTCTCTATTTGTTT 2401
Qy 762 LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys 781
Db 2402 CTAGTCTTGGAAATATTTCTTTTATATATTTTCTCCAGGGCTTCACATTTGGCAAA 2461
Qy 782 AlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGln 801
Db 2462 GCTGGGAGATCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTGAGACAG 2521
Qy 802 AspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla 821
Db 2522 GATGTCAGCTGGTGTGATGACCCCTAAAAACACCACTGGAGCATTGACAAACCGAGCTTGC 2581
Qy 822 AspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn 841
Db 2582 AATGATCGCGCTCAAGTTAAAGGGCTATAGGTTCAGCGCTTGTCTCAATACCCAGAAAT 2641
Qy 842 IleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeu 861
Db 2642 ATAGCAAACTCTGGGACAGCATTTATATCTTAATCTATGTTGGCAATTAACACTT 2701
Qy 862 LeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLysMetLeu 881
Db 2702 TTACTCTTAGCAATTTGACCAATCATTTGCAATAGCAGGAGTTGTTGAAATGAAATGTTG 2761
Qy 882 SerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGlu 901
Db 2762 TCTGGACAGCACTGAAGATAGAAAGAGCTAGAGGAGCTGGGAAGATTCCTACAGAA 2821
Qy 902 AlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyrMet 921
Db 2822 GCCATCGAAACCTCCGAACCTGTTGTTTCTTGTGACTCGGGACGACGAAGTTTGAATACATG 2881
Qy 922 TyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGly 941
|||||

Db 2882 TATGCACAGAGTTTGCAAAGTACCATACAGAAACTCTTTTGAGAAAGCACACATCTTCGGG 2941
Qy 942 ValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPhe 961
Db 2942 GTCTCATTTTCTATCACCCAGCAATGATGATATTTTCTCTATGCTGGCTGTTCCGGGTTT 3001
Qy 962 GlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSer 981
Db 3002 GGTGCTACTTGGTGGCAATGAGTTCAATGAATCTTCCAGGATGTTCTTTTGGTATCTCA 3061
Qy 982 AlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyrAla 1001
Db 3062 GCTATTTGCTTTGGTGGCCATGGCAGTGGGCGAGTCACTTTCATTTGCTCTGAGCATATGCC 3121
Qy 1002 LysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIleAsp 1021
Db 3122 AAAGCAAGATATCAGACGCCACGTCATCATGATCATTTGAAAAAGCCCTCTGATTGAC 3181
Qy 1022 SerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGlu 1041
Db 3182 AGCTACAGCCCTCAGGGCTCAAGCCAAATACGTTGGAAGGAAATGTGACATTTAATGAG 3241
Qy 1042 ValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeuGlu 1061
Db 3242 GTCTGTTTCAACTATCCCACTCGACACATCCCGTCTCCAGGGGCTGAGCCTCGAG 3301
Qy 1062 VallLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrVal 1081
Db 3302 GTGAAGAAGGCCAGACGCTGGCCCTCTGTAGTAGCAGTGGCTGTGGAGAGACACAGTT 3361
Qy 1082 ValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGlyLys 1101
Db 3362 GTTCAGCTCTTAGAGCGCTTCTATGACCCCTTGCTGCTGCTAGTGTAAATGATGGCAAA 3421
Qy 1102 GluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGlu 1121
Db 3422 GAGATAAAGCACCTGGAATGTCCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAGGAG 3481
Qy 1122 ProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgVal 1141
Db 3482 CCCATCTGTTTGTACTGCAGCATTTGCCGAGAACATTTGCCATGGAGACACAGCCGGGTC 3541
Qy 1142 ValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisPheIleGlu 1161
Db 3542 GTATCACATGAAGAGATTATGAGGAGCCCAAGGAGGCCAACATACACACCATCTTCATCGAG 3601
Qy 1162 ThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGly 1181
Db 3602 ACATCTCCCTGAGAATACACACCAGAGTAGGAGACAAAGAACCCAGCTCTCTGTGGTGGC 3661
Qy 1182 GlnLysGlnArgIleAlaIleAlaAlaLeuValArgGlnProHisIleLeuLeuLeu 1201
Db 3662 CAGAAACAGCCGATTTGCCATAGCTCGCGCTCTTCTTAGACAGCCCTCATATTTGCTTTG 3721
Qy 1202 AspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeuAsp 1221
Db 3722 GATGAAGCTACATCAGCTCTGGATACAGAAAGTCAAAAGGTTGTCCAAGAGCCCTGGAC 3781
Qy 1222 LysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsn 1241
Db 3782 AAAGCAGAGAAAGCCGACCTGCTGATGATGCCACCCGCTTGTCCACCATCCAGAAAT 3841
Qy 1242 AlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGlnGln 1261
Db 3842 GCATATTTAATAGTGGTGTTCAGAAATGGCAAGTCAAGGAGCATGGCACACATCAACAG 3901
Qy 1262 LeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLysArg 1281
Db 3902 CTGCTGGCTCAGAAAGGCACTATTTTTCATGATCAGTGTCCAGGCTGGAGCAAAACGC 3961
RESULT 6
AF045016 4317 bp mRNA linear MAM 07-FEB-1998
LOCUS

DEFINITION Canis familiaris multidrug resistance p-glycoprotein (MDR1) mRNA, complete cds.

ACCESSION AF045016

VERSION AF045016.1 GI:2852440

KEYWORDS

SOURCE dog.

ORGANISM Canis familiaris

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

AUTHORS 1 (bases 1 to 4317)

TITLE Puel,O., Lepage,J.F., Alvinerie,M., Galtier,P. and Pineau,T.

JOURNAL Direct Submission

Submitted (28-JAN-1998) Pharmacology, INRA, BP 3, 180 Chemin de Tournefeuille, Toulouse Cedex 9 31931, France

FEATURES

source location/Qualifiers

1..4317

/organism="Canis familiaris"

/db_xref="taxon:9615"

/tissue_type="colon"

/note="beagle"

1..4317

/gene="MDR1"

70..3912

/gene="MDR1"

/codon_start=1

/product="multidrug resistance p-glycoprotein"

/protein_id="AAC02113.1"

/db_xref="GI:2852441"

/translation="MDPEGRKGRGAEKFNWKMKGKSKKKEKKKPTVSTFAMFRYSNW LDRLYMLVGTMAIIGHAALPLMLVFNMTDSFANAGISRNKTTPVPIINESITNNQ HFNLHEEMTVAYYSIGAGVLVAAYIOVSWCLAAAGROILKIRKOFFHAILMROE IGFWDVHDGELNTRITDDVSKINEGIGDKVMGTOSIATFTGTGTGTGPKWKLTLV ILAISVPLGSAIAWAKILSSFTDKELLAYAKAGAAVEVLAARIVAFGQKKELE RYNNLEAKRIGIKKAITANISIGAFLIYALAFWIGTSLVSISYETIGQVLT VFPSVLGAFSGIQASPSIEAFANARGAAEYFIKIIDNKPISDSYKSGHKPDNIGN LEPKNHFYSPSRKILKLNKQSGQTVALVNGSGCKSTVQLMQRLYDPTDG MWCIDGQDRTINVRHLREITGVVSOEPLVFAITTAENIRYGRNVMTDEIEKAVKEA NAYDFIMKLPNKDTLVGERGLSGQKORIAIARLVNPKILLDLDEATSDALPTES EAVQVALDKARKGRTTIVIAHLSVRNADVIAGFDGQVIYEKGNHDELMKEGIYF KLVMTQTRNELENAESKESSEDALEMSPDSSGLIKRKRSTRSRSHAFQGGQRK LGPKEDLNENPVSFWRIKLKNSTENPVVVGIFCAIINGGLQPAFISIFRIIGIF TRDEPDKTRQNSMF SVLFLVLIISFTFFLQGTFGKAGEIITKRLRYVFRSML RQVSDVPDPKNTGALTTRLANDAAQVKAIGSLAVITQNIANLGTGIIISLIYGW QLPFLALLAPIIAIGVVMKMSQALKDKKELEGAGKIATEAENRTVVSLETR EOKFYWASLOVPMYRNSLRKAHIFGVSPSITCAMYFVACFRGAYLVANEFNFE QDVLLVFSIAVFGMAVQGVSPADYAKAKVSAAHVIMIEKSPILDSYSPHGLKPN TLEGNVTFNWVFNPTRPDIPVLQGLSLEVKQFTLALVSGSGCKSTVQVLLERFY DPLAGSVLIDGKEIKHLNWOLRAHLGIYSQEPILFDCSIAENIAYGDNRSVSHEEI MQAAKEANIHHFTETLPKYNTRVKGDTQLSGGQKORIAIARLVQRPKHILLDEAT SALDTESEKVVQBALDKAREGRTCIIVIAHRLSTIQNALIIVVFQNGKVEHGHQQL AOKGIYFMSISVQAKR"

BASE COUNT 1293 a 844 c 1019 g 1161 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4317

Score: 6425.50 Matches: 1273

Percent Similarity: 99.69% Conservative: 4

Best Local Similarity: 99.38% Mismatches: 3

Query Match: 99.25% Indels: 1

DB: 4 Gaps: 1

US-09-672-725C-23 (1-1281) x AF045016 (1-4317)

Qy 1 MetAspProGluGlyArgGlySerAlaGluLysAsnPhetTrpLysMetGlyLys 20

|||||

Db 70 ATGGATCTGAAGGAGCGGTAAAGGAGTGCAGAGAAGAACTTCTGGAAATGGGCAA 129

Qy 21 LysSerLysLysAsnGluLysGluLysLysProThrValSerThrPheAlaMetPhe 40

|||||

Db 130 AAAAGTAAAAA---GAGAGAGAAAGAAACCACTGTGCAGACGTTTGCATATGTT 186

Qy 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 60

|||||

Db 187 CCCTATTCAAATGGCTTGATAGTTGT ||||| "TGCCATCATC 246

Qy 61 HisGlyAlaAlaLeuProLeuMetMetL spSerPheAla 80

|||||

Db 247 CATGGAGCTGCACCTCCCTCTCATGATGC ATAGCTTTGCA 306

Qy 81 AsnAlaGlyIleSerArgAsnLysThrP erIleThrAsn 100

|||||

Db 307 AATGAGGAATTTCAAGAAACAAAACCTT GTATTACGAC 366

Qy 101 AsnThrGlnHisPheIleAsnHisLeu latTyrTyr 120

|||||

Db 367 AATACACAACATTCATCAACCATCTGGAGGAGGAAATGACCACGTATGCTATTATTAC 426

Qy 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140

|||||

Db 427 AGTGGATCGGTGCTGGCGTCTGTGTGGCTTACATCCAGGTTTCATCTCGGTGCTG 486

Qy 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160

|||||

Db 487 GCAGCAGGAGACAGATCTCAAAATTAGAAAACAAATTTTTCATGCTATCATGCGACAG 546

Qy 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180

|||||

Db 547 GAGATTGGCTGTTTGACCTGCATGACGTGGGAGCTTAACACCGCGCTCACAGACGAT 606

Qy 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200

|||||

Db 607 GTCTCCAAATCAATGAAGAAATTCGCGACAAGTTGGAATGTTCTTCAATCAATAGCA 666

Qy 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220

|||||

Db 667 ACATTTTTCACCGTATTATAGTGGGTTCACCTGGTGGAGCTAACCTTGTGATT 726

Qy 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240

|||||

Db 727 TTGGCCATCAGCCCTCTTCTGACCTTTCAGCGCCATCTGGGCAAGATATCTCTTCA 786

Qy 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu 260

|||||

Db 787 TTTACTGATAAGAACTCTTGGCCTATGCAAGCTGGAGCAGTAGCTAGCAAGAGCTT 846

Qy 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280

|||||

Db 847 GCAGCAATCAGACTGTGATTGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAC 906

Qy 281 LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300

|||||

Db 907 AAAAATTTAGAGAAGCTAAAAGAAATGGGATAAGAAAGCTATCACGCCCAACATTCT 966

Qy 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320

|||||

Db 967 ATTGGTGGCGCTTCTTATTGATCATGATCATATGCTCTGGCTTCTGGTATGGGACC 1026

Qy 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340

|||||

Db 1027 TCCTTGGTCTCTCCAGTGAATATACTATTGGACAGGTACTCACTGCTCTCTTCTTCTGA 1086

Qy 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360

|||||

Db 1087 TTAATTGGGGCTTTTATATTGGACAGCATCCCAAGCATTTGAAGCATTTGCCAACGCA 1146

Qy 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380

|||||

Db 1147 AGAGGAGCAGCTTATGAATCTTCAAGATAATTTGACAAATAACCAAGCATTTGACAGTAT 1206

Qy 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400

|||||

Db 1207 TCGAAGAGTGGACATAAACACAGATATATTAAGGGAAATTTGGAATTTCAAAATGTTTAC 1266

Qy 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420

|||||


```
QY 1141 ValValSerHisGluGluMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1160
Db 3487 GTGATACATGAGAGATTATGAGGAGCCAGGAGCCCAACATACACCATTCATC 3546
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
Db 3547 GAGACACTCCCTCAGAAATACACACAGAGTAGGAGACAAAGAACCCAGCTCTCTGT 3606
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
Db 3607 GGCAGAAACAGCGCATTCGATAGCTCGCGCTCTTGTAGACAGCCTCATATTTTGT 3666
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
Db 3667 TTGATGAGACCTACATCAGCTCTGGATACAGAAAGTGAAGGTTGTCCAAGAGCCCTG 3726
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
Db 3727 GACAAAGCCAGAGAGCCGACCTGCATTGTGATCGCCACCGCTTGCCACCATCCAG 3786
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
Db 3787 AATGCAGATTAAATAGTGGTGTTCAGATGGCAAGTGCAGAAAGTCAAGAGAGCATGCCACATCAA 3846
QY 1261 GlnLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
Db 3847 CAGCTGCTGCTCAGAAAGCATCTATTTTCCATGATCATGCTCCAGGCTGGAGCAAG 3906
QY 1281 Arg 1281
Db 3907 CGC 3909

RESULT 7
AX105059
LOCUS AX105059 4317 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 3 from Patent WO0123540.
ACCESSION AX105059
VERSION AX105059.1 GI:13921211
KEYWORDS
SOURCE
  ORGANISM
    Canis familiaris
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
  REFERENCE
    1 (bases 1 to 4317)
    Stocker,P.J., Steimel-Crespi,D.T., Crespi,C.U., Reif,T.C. and
    Patten,C.J.
  TITLE P-glycoproteins and uses thereof
  JOURNAL Patent: WO 0123540-A 3 05-APR-2001;
  GENTEST CORPORATION (US)
  FEATURES
    source
      1..4317
        /organism="Canis familiaris"
        /db_xref="taxon:9615"
        /note="unnamed protein product"
        /codon_start=1
        /protein_id="CAC37719.1"
        /db_xref="GI:13921212"
        /translation="MDPEGRKGSAEKNFWMGKSKKSKKKKPTVTFAMFRYSNW
        LDRYLMVGTMAIIRGAALPLMLVFGNMTDSFANAGISRNKTPFVILNESITNNQ
        HFNHLEEMTVAIYYSIGIGAGLVAAVIOVSFWCLAAAGRQILKIRKQFFHAIKRQ
        IGWFDVHVGELNTRLTDVSKINEGIGKGVGMFOSIATFTTGTGKTVPGWKLTLV
        ILAISVGLPSAIWAKILSSFTDKELLAVAKAGAAVEELAAIRTVIAFGOKKELE
        RYNNLEAKRIGIKKAITANISIGAIFLLIYASALAFWYGTSLVLSSEYTIQVLT
        VFSVLIFAFSISQASPSIEAFANARGAAEYFIKIDNPSIDSYSKSHGKPDNKG
        LEFNHYFSYPSKEVKILGLNKVQSGTVALVNGSCGKSTTVQLMORLYDPDG
        MVDIGDQIRTNVRLREITGVVSOEPLVFATIAENIRYGRNVMTDEIEAKVEA
        NAYDFTMKLPNKFDTLVBGERLARGSGOKORIAARALVRNPKILLDEATSAIDTES
        FAVOVALDKARKGRTIVIAHRLSTVRNADVDFDGVIVIEKGNHDELMKEKGIYF
        KLYMTQRCNEIELENATGESKSALEMSPKDSGLIKRSTRSTRSIAHQGDRIK
        LGTKEDLNENVPSFWRILKNSDEWPYFVGIFCAIINGGLQPAFSIIFSKIIGIF
        TRDEPVSFKRQNSNMFVSLFVLGIISFTFFLQGTFTGKAGEILTKRLRYWVFRSML
        RODVSFWFDDPKNTTGTALTTLRLANDAQAQVKAIGSRLAVITQNIANLGTIILSIYWG
```

```
QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 70 ATGGATCTCTGAAGGAGCGGTAAGGGAGTGCAGAGAAGAACTCTGGAAAATGGGCAA 129
QY 21 LysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
Db 130 AAAAGTAAAAA--GAGAAAGAAAGAAACCAACTGTCCAGCAGCTTTGCAATGTT 186
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 60
Db 187 CGCTATTCAATGGCTTGATAGTTGTATATGTTGGTGGGACAAATGGCTCCATCATC 246
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
Db 247 CATGGAGCTGCACCTCCCTCATGATGCTGGTGTGGTGGGAAACATGACAGATAGCTTGA 306
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 307 AATGCAGGAATTCAAGAAACAAACTTTCCAGTTATATAATTAATGAAAGTATTACGAAC 366
QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 120
Db 367 AATACACAACATTTTCATCAACCATCTGGAGAGGAAATGACCACCATGCTGCTATTATTAC 426
QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
Db 427 AGTGGGATCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 486
QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
Db 487 GCAGCAGGAGAGACAGATCTCAAAATTAAGAAACAAATTTTTCATGCTATCATGCGACAG 546
QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
Db 547 GAGATTGGCTGGTTTGACGTGATGACGTGGGGAGCTTAAACACCGGCTCACAGACCAT 606
QY 181 ValSerLysLysLeuGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
Db 607 GTCTCAAAATCAATGAAGAAATGGGACAAAGTTGGAAATGTTCTTCAATCAATAGCA 666
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
Db 667 ACATTTTTCACCGGTTTTATAGTGGGGTTTACACCTGGTTGGAAGCTAACCCCTTGTA 726
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
Db 727 TTGGCCATCAGCCCTCTCTTCTGACCTTTTCAGCCGCATCTGGGCAAGATATATCTTCA 786
QY 241 PheThrAspLysGlnLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu 260
Db 787 TTTACTGATAAAGAACTCTTGGCCTATGCAAAAGCTGGAGAGTAGTACCTGAAGAGCTCTA 846
```

```
QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 70 ATGGATCTCTGAAGGAGCGGTAAGGGAGTGCAGAGAAGAACTCTGGAAAATGGGCAA 129
QY 21 LysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
Db 130 AAAAGTAAAAA--GAGAAAGAAAGAAACCAACTGTCCAGCAGCTTTGCAATGTT 186
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 60
Db 187 CGCTATTCAATGGCTTGATAGTTGTATATGTTGGTGGGACAAATGGCTCCATCATC 246
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
Db 247 CATGGAGCTGCACCTCCCTCATGATGCTGGTGTGGTGGGAAACATGACAGATAGCTTGA 306
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 307 AATGCAGGAATTCAAGAAACAAACTTTCCAGTTATATAATTAATGAAAGTATTACGAAC 366
QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 120
Db 367 AATACACAACATTTTCATCAACCATCTGGAGAGGAAATGACCACCATGCTGCTATTATTAC 426
QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
Db 427 AGTGGGATCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 486
QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
Db 487 GCAGCAGGAGAGACAGATCTCAAAATTAAGAAACAAATTTTTCATGCTATCATGCGACAG 546
QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
Db 547 GAGATTGGCTGGTTTGACGTGATGACGTGGGGAGCTTAAACACCGGCTCACAGACCAT 606
QY 181 ValSerLysLysLeuGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
Db 607 GTCTCAAAATCAATGAAGAAATGGGACAAAGTTGGAAATGTTCTTCAATCAATAGCA 666
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
Db 667 ACATTTTTCACCGGTTTTATAGTGGGGTTTACACCTGGTTGGAAGCTAACCCCTTGTA 726
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
Db 727 TTGGCCATCAGCCCTCTCTTCTGACCTTTTCAGCCGCATCTGGGCAAGATATATCTTCA 786
QY 241 PheThrAspLysGlnLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu 260
Db 787 TTTACTGATAAAGAACTCTTGGCCTATGCAAAAGCTGGAGAGTAGTACCTGAAGAGCTCTA 846
```

BASE COUNT 1293 a . 844 c 1019 g 1161 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4317
Score: 6425.50 Matches: 1273
Percent Similarity: 99.69% Conservative: 4
Best Local Similarity: 99.38% Mismatches: 3
Query Match: 99.25% Indels: 1
DB: 6 Gaps: 1

US-09-672-725C-23 (1-1281) x AX105059 (1-4317)

Qy	261	AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn	280
Ile			
Db	847	GCAGCAATCAAGACTGTGATTGGCTTTGGAGCAGCAAAAGAAAGAACTTTGAAAGGTACAAC	906
Qy	281	LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer	300
Ile			
Db	907	AAAAATTTAGAAGAGCTAAAGAAATTTGGGATAAAGAAAGCTATCACGCCCAACATTTCT	966
Qy	301	IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr	320
Ile			
Db	967	ATTGGTGGCGCTTCTATTATGATCATGCATCATATGCTCTGGCTTCTTGGTATGGGACC	1026
Qy	321	SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal	340
Ile			
Db	1027	TCCTTTGGTCTCTCCAGTGAATATATATTATGGCAGGTACTCACGTCTCTTTCTTCTGTA	1086
Qy	341	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	360
Ile			
Db	1087	TTAATTTGGGCTTTTAGTATTGGACAGGCATCCCAAGCATTTGAAGCATTTTGCAACGC	1146
Qy	361	ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	380
Ile			
Db	1147	AGAGGACAGCTTATGAAATCTTCAAGATAATTGACATAAACCAAGCATTCACAGCTAT	1206
Qy	381	SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	400
Ile			
Db	1207	TCGAAGAGTGGACATAAACAGATAATAATTAAGGGAAATTTGGAAATTCAAAAATGTTCC	1266
Qy	401	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	420
Ile			
Db	1267	TTCAAGTTACCCCTTCTCGAAAGAAGTTAAGATCTTTAAAGGGTCTCAACTGAAGGTTC	1326
Qy	421	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrValGln	440
Ile			
Db	1327	AGTGGGACAGACTGGCGCTGGTTGGACACATGGCTGCGGGAGAGCAGCACCGTCGAC	1386
Qy	441	LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	460
Ile			
Db	1387	CTGATGCAGAGGCTCTATGCCCCACAGATGCATGGTCTGTATTGATGGACAGACATT	1446
Qy	461	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	480
Ile			
Db	1447	AGGACCATAAATTAAGGCATCTTCGGGAAATTAAGTGTGGTGGTGCAGTGCAGGACCTGTG	1506
Qy	481	LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	500
Ile			
Db	1507	TTGTTTCCACACAGATGCTGAAACATTCGCTATGCCCGCGGAAATATGTCACCATGGAT	1566
Qy	501	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	520
Ile			
Db	1567	GAGATTGAGAAAGCTGTTAAGGAAGCAATGGCTATGATTTTATCATGAAACTACCTAAT	1626
Qy	521	LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg	540
Ile			
Db	1627	AAATTTGACACTCTGGTTGGAGAGAGGGGGCCCGGTGAGTGTGGCAGAGAACAGAGA	1686
Qy	541	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	560
Ile			
Db	1687	ATCGCCATTCTCGGGCCCTGGTTCCGAACCCCAAGATCTCTTCGCTGGATGAGGCAACG	1746
Qy	561	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	580
Ile			
Db	1747	TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAAGTGGCCCTGGATAAGGCCAGAAA	1806
Qy	581	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	600
Ile			
Db	1807	GGCCGGACATACCATTGTGATAGCTCATCGTTTGGTCTACAGTTCGTAATGCGGATGCTATT	1866
Qy	601	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	620
Ile			
Db	1867	GCTGTTTTCATGATGCGACTCATCTTGGAGAAGGAATCATGATGAACCTCATGAAGAG	1926
Qy	621	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	640

Db	1927	AAGGCAATTTACTTCAAACTTGTCACAAATGCACAAAGAGGAAATGAAATTCAGTGTAGAA	1986
Qy	641	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	660
Db	1987	ATNGCCACTGGTGAATCCAAAGTGAAAGTGATGCCTTGGAAATGCTCCAAAGATTCA	2046
Qy	661	GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln	680
Db	2047	GGGTCCAGTTTAAATATAAAAGAAAGTCAACTCGCAGAGTATACATGCACCAAGGCCAA	2106
Qy	681	AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp	700
Db	2107	GACGAARACCTTGGTACAAAGAGGACCTGGAATGAGAATGTTCCTTCAGTTTCCCTTCGG	2166
Qy	701	ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla	720
Db	2167	AGGATTCGAAAGCTGAACCTCAACTGAATGGCCCTATTATTGTGGTGGTATATTTTGTGCT	2226
Qy	721	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	740
Db	2227	ATTATAACAGGAGCCCTGCACACAGCATTTTCAATAATATTTTCAAGGATTATAGGATC	2286
Qy	741	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	760
Db	2287	TTTACCCGAGATGAGGATCCTGAAACAAACGACAGAAATAGTAACATGTTTCTGTATTG	2346
Qy	761	PheLeuValLeuGlyIleIleIleSerPheIleThrPhePheLeuGlnGlnGlyPheThrPheGly	780
Db	2347	TTTCTAGTCCCTGGAAATATTTCTTTTATACATTTTCTCCAGGGCTTCACATTTTGGC	2406
Qy	781	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	800
Db	2407	AAAGCTGGGAGATCCCTCACTAAGCGCTTCGATACATGGTTTTCAGATCCATCCTCAGA	2466
Qy	801	GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	820
Db	2467	CAGATGTCAGCTGGTTGTGATGACCTTAAAAACACACACTGGAGCATTCACAAACAGCCT	2526
Qy	821	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	840
Db	2527	GCCAAATGATGGGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTGCTGTCATTTACCAG	2586
Qy	841	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr	860
Db	2587	AATATAGCAAAATCTGGACAGGCATTTATATATCTTAATCTATGTGTGGCAATTAACA	2646
Qy	861	LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet	880
Db	2647	CTTTTACTCTTAGCAATGTACCCTCATGTGCATAGCAGAGTTGTGAAATGAAATG	2706
Qy	881	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGlyAlaGlyLysIleAlaThr	900
Db	2707	TTGTCTGCACAAGCACTGAAAGATAAGAAAGAGCTAGAAGGAGCTGGAGATTGCTTACA	2766
Qy	901	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr	920
Db	2767	GAAGCCATCGAAACATTCGCAACTGTGTGTTCTTTGACTCGGAGACAGAGTTTGAATC	2826
Qy	921	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	940
Db	2827	ATGTTATCCAGAGATTGGCAATACCATACAGAAACTCTTTGAGGAAAGCACACATCTTC	2886
Qy	941	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	960
Db	2887	GGGTCTCATTTTCTATCACCAGGCAATGATGATTTTCTCTATGCTGCTGCTTCCGG	2946
Qy	961	PheGlyAlaTrpLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe	980
Db	2947	TTTGGTCCCTACTTGTGTGGCAATGAGTTCATGAACCTTTCAGGAGTCTTTTGGTATTC	3006
Qy	981	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	1000

Db	3007	TCAGCTATTGCTTGTGGTCCAGTCGGCAGTCAGTTCATTGCTCCTGACTAT	3066
Qy	1001	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1020
Db	3067	GCCAAAGCCAAAGATATCAGCAGCCAGCTCATCATCATTTGAAAAAGCCCTCTGATT	3126
Qy	1021	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1040
Db	3127	GACAGCTACAGCCCTCAGGCCTCAGCCAAATACGTGGAAGAAATGTGACATTAAAT	3186
Qy	1041	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1060
Db	3187	GAGTCTGTTCACCTATCCACTCGACACAGACATCCCGTGTCCAGGGGCTGAGCCTC	3246
Qy	1061	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1080
Db	3247	GAGGTGAAGAGGGCCAGAGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGACACA	3306
Qy	1081	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1100
Db	3307	GTGTTTCAGCTCCTAGAGCCGCTTCTATGACCCCTTGGCTGGTTCAGTGCTTAATGTAGGC	3366
Qy	1101	LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1120
Db	3367	AAAGAGATAAGCACCTGAATGCCAGTGGCTCGGACACACCTGGGCATCGTGTCTCAG	3426
Qy	1121	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1140
Db	3427	GAGCCCATCTGTTGACTGTCAGCATTCGCGAGAACATTCGCTATGGAGACAACAGCGG	3486
Qy	1141	ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle	1160
Db	3487	GTCTATACATGAAGAGATTATGCAGGAGCCCAAGGAGGCCAACATACACCATTCATC	3546
Qy	1161	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1180
Db	3547	GAGACATCCCTGAGAAATACACACCCAGAGTAGGACACAAGGAAACCCAGCTCTCGT	3606
Qy	1181	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1200
Db	3607	GGCCAGAAACAGCGCATGTCATAGCTCGCGCTCTTGTAGACAGCGCTCATATTTGCTT	3666
Qy	1201	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1220
Db	3667	TTGGATGAGCTACATCAGCTCTGGATACAGAAAGTGTGTCGAAGAGCCCTG	3726
Qy	1221	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1240
Db	3727	GACAAAGCCAGAGAAGCCGACCTGCATTGTGATCGCCACCGCTTGTCCACCATCCAG	3786
Qy	1241	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1260
Db	3787	AATCAGATTAAATAGTGGTGTTCAGAAATGGCAAGTCAAGGAGCATGGCACACATCAA	3846
Qy	1261	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys	1280
Db	3847	CAGTGTGGCTCAGAAAGGCATCTATTTTCCATGATCAGTGTCCAGGCTGGAGCAAG	3906
Qy	1281	Arg	1281
Db	3907	CGC	3909
RESULT 8			
CFA419568			
LOCUS			
DEFINITION			
Canis familiaris mRNA for abortive P-glycoprotein (p-gp gene).			
ACCESSION			
AJ419568			
VERSION			
AJ419568.1 GI:17385398			
KEYWORDS			
P-glycoprotein; p-gp gene.			
SOURCE			
dog.			
ORGANISM			
Canis familiaris			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
REFERENCE	1	(sites)	
AUTHORS	Roulet A.		
TITLE	Characterization of a MDR1a P-gp deficient dogin relation to		
JOURNAL	ivermectin sensitivity		
REFERENCE	2	(bases 1 to 3934)	
AUTHORS	Roulet A.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-NOV-2001) Roulet A., Pharmacology, Inra, 180 Chemin		
FEATURES	de Tournefeuille, 31931 Toulouse cedex, FRANCE		
source	Location/Qualifiers		
gene	1..3934		
CDS	/organism="Canis familiaris"		
	/db_xref="taxon:9615"		
	/tissue_type="liver"		
	/note="ivermectin sensitive collies"		
	7..282		
	/gene="p-gp"		
	7..282		
	/gene="p-gp"		
	/note="abortive product"		
	/codon_start=1		
	/evidence="experimental"		
	/product="p-glycoprotein"		
	/protein_id="CAD12806.1"		
	/db_xref="GI:17385399"		
	/translation="MDPEGRRGSAENKFWKMGKSKKNEKKKPKPVSTFAMFRYSN		
	WLDRLYMLVGTMAAIIHGAAALPLMLLVGNMTALQMOFQETKLFLQ"		
BASE COUNT	1144 a	815 c	964 g
ORIGIN	1011 t		
Alignment Scores:			
Pred. No.:	0	Length:	3934
Score:	6416.50	Matches:	1274
Percent Similarity:	99.61%	Conservative:	2
Best Local Similarity:	99.45%	Mismatches:	4
Query Match:	99.11%	Indels:	2
DB:	4	Gaps:	1
US-09-672-725c-23 (1-1281) x CFA419568 (1-3934)			
Qy	1	MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys	20
Db	7	ATGATCTCTGAGGAGCGCTAGGGGAGTGCAGAGAAGAACTTCTGGAAGATGGCAAA	66
Qy	21	LysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe	40
Db	67	AAAAAGTAAAAAATGAGAAAGAAAGAAACCAACTGTCCAGCAGTTTGCAATGTTT	126
Qy	41	ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle	60
Db	127	CGCTATTCAATGGCTTGTATGTTGTTGGTGGGACAAATGGCTGCCATCATC	186
Qy	61	HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	80
Db	187	CATGGAGTGCACCTCCCTCATGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	242
Qy	81	AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	100
Db	243	AATGCAGGAATTTCAAGAAACAAACTTTTCCAGTTTATTAATTAAGAAAGTATTACGAAC	302
Qy	101	AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr	120
Db	303	AATACACACATTTTCAACCACTCTGGAGGAGGAAGTAATGACCGTATGCCATTATTATAC	362
Qy	121	SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu	140
Db	363	AGTGGGATCGGTGCTGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	422
Qy	141	AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln	160
Db	423	GCAGCAGGAGACAGATCTCAAAATAGAAACAAATTTTTTCATGCTATCATGCGACAG	482

|||||
Db 2643 TTGTCTGGCAAGCACTCAAGATAGAAAGAGCTAGAGGAGCTGGGAAGATGCTACA 2702
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2703 GAAGGCATCGAAACTCCGAAGTCTTTCTTTGACTCGGAGCAGAACTTTGAATAC 2762
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2763 ATGTATGCACAGATTTCAGAGTACCATACAGAACTCTTTGAGGAAGCACACATCTTC 2822
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2823 GGGGTCTCTATTCATCCAGCAATGATATTTTCTCTATGCTGGCTGTTCCGG 2882
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
Db 2883 TTTGGTGCTACTTGGTGGCAATGAATTCATGAACTTTCAGGATGTTCTTTGGTATTC 2942
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
Db 2943 TCAGCTATTGCTTTGGTGCCATGCGACTGGGCGAGTCACTTCACTTGTCTCTGACTAT 3002
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
Db 3003 GCCAAGCCAAAGTATCAGCAGCCACGTCATCATGATCATGAAAAAGCCCTCTGATT 3062
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3063 GACAGCTACAGCCCTCAGGGCTCAAGCCAAATACGTTGGAAGAAATGTGACATTTAAT 3122
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3123 GAGGTCGTGTTCACTATCCACTCGACCACACATCCCGTGCCTCCAGGGCTCAGCCTC 3182
QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
Db 3183 GAGGTGAAGAGGGCGACAGCTGCCCTCGTAGTAGCAGTGGCTGTGGGAAGAGCACCA 3242
QY 1081 ValValGlnLeuLeuGluAargPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
Db 3243 GTTGTTCAGCTCCTAGAGCGCTTCTATGACCCCTTGGCTGGTTCAGTGCTAAATGTATGGC 3302
QY 1101 LysGluIleLysHisLeuAsnValGlnTyrLeuArgAlaHisLeuGlyIleValSerGln 1120
Db 3303 AAAGAGATAAGACCTGGAATGTCCAGTGGCTCCGAGCACACCTTGGGCATCGTCTCAG 3362
QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
Db 3363 GAGCCCATCTCTTTGACTGCAGCATTCGCCGAGAACATTGGCTATGGAGACAACAGCCGG 3422
QY 1141 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1160
Db 3423 CTCCTATACATGATGAAGATTTATCAGCGACGCCAAGAGGCGCAACATACACACTTCATC 3482
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
Db 3483 GAGACACTCCCTGAGAATACAAACCCAGAGTAGGAGACAAGAACCCAGCTCTCTGGT 3542
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaAargAlaLeuValArgGlnProHisIleLeuLeu 1200
Db 3543 GGCCAGAAACAGCGATTGCCATAGCTCGCGCTCTTGTGTAGACAGCTCATATTTTCTT 3602
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
Db 3603 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGTCCAAAGAGCCCTG 3662
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
Db 3663 GACAAAGCCAGAGAGCCGACCTGCTATTGTATGCCCCACCCGCTTGTCCACCATCCAG 3722
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
|||||

Db 3723 AATGCAGATTAAATAGTGGTGTTCAGAAATGCAAAAGTCAAGGAGCATGGCACACATCAA 3782
QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
Db 3783 CAGCTGTGCTGCAGAAAGCATCTATTTTCCATGATCAGTGTCCAGGCTGGAGCAAA 3842
QY 1281 Arg 1281
|||
Db 3843 CGC 3845
RESULT 9
AX322787
LOCUS AX322787 3860 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 1 from Patent WO0192877.
ACCESSION AX322787
VERSION AX322787.1 GI:18093766
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Sorrentino,B. and Schuetz,J.
TITLE Method of identifying and/or isolating stem cells
JOURNAL Patent: WO 0192877-A 1 06-DEC-2001;
ST. JUDE CHILDREN'S RESEARCH HOSPITAL (US)
FEATURES
source Location/Qualifiers
I. :3860
BASE COUNT 1135 a 746 c 957 g 1022 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3860
Score: 5866.00 Matches: 1164
Percent Similarity: 95.09% Conservative: 55
Best Local Similarity: 90.80% Mismatches: 59
Query Match: 90.61% Indels: 4
DB: 6 Gaps: 4
US-09-672-725c-23 (1-1281) x AX322787 (1-3860)
QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheThrLysMetGly 19
Db 1 ATGGATCTTGAAGGGGACCCCAATGAGGAGCAAGAAAGAACTTTTAAACTGAAC 60
QY 20 LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 61 AATAAAGTGAAAA--GATAAGAGGAAAGAACCAACTGTCAAGTGTATTTTCAATG 117
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
Db 118 TTTTCGCTATTCAAAATTTGGCTTGACAAAGTTGATATGTTGGTGGAACTTTGGCTGCATC 177
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 178 ATCCATGGGGCTGGAGCTTCTCTCATGATGCTGGTGTGGAGAAATACAGATATCTTT 237
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 238 GCAATGCAGGA--AATTTAGAAGATCTCATGTCAACATCACTAATAGAAGTGATATC 294
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 119
Db 295 AATGATACAGGGTCTTCATGAAT---CTGGAGGAAGACATGACCATATGCTATTAT 351
QY 120 TyrSerGlyIleGlyAlaGlyValLeuAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 352 TACAGTGAATATGGTCTGGGGTGGTGGTGTGCTGTACATTACAGGTTTCATTTGGTGC 411
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
|||||

Db 412 CTGGCAGCTGGAAGACAAATACACAAAATAGAAAACAGTTTTTTCATGCTATAATGGCA 471
QY 160 GlncluleglyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db CAGAGATAGGCTGGTTGATGTGCACAGATGTGGGAGCTTAACACCGACTTACAGAT 531
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db GATGTCTCTAAGATTAAATGAAGGTATTGTGTGACAAAATTTGGAATGTCTTTTCAGTCAATG 591
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db GCACCAATTTTCACTGGGTTTATAGTAGGATTACACGTGGTGGAGGCTAACCCCTGTG 651
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db ATTTTGGCCCATCAGTCTCTGTCTTGACCTGTCTGACCTGTCTGGCCAAAGATACATCT 711
QY 240 SerPheThrAspLysGluLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db TCATTTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 771
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysGluLeuGluArgTyr 279
Db TTGGCCACCAATTAGAACTGTGATTGCTATTTGGAGGACAAAGAAAGAACTTGAAGGTTAC 831
QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db AACAAAAATTTAGACAGACTAAAGAAATTTGGATTAAGAAAGCTATTACACCCATATT 891
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db TCTATAGGTCTCTCTTCCTGCTGATCTATGCTATCTGCTGCTGCTGCTGCTGCTGCTGCT 951
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db ACCACCTTGGTCTCTCAGGGGAATATTCTATTGGACAACTACTCCTACTATTCTTTCTT 1011
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db GTATTAAATTTGGGCTTTTAGTGTGGACAGGCATCTCCAAGCATTTGAACCAITTTGCAAT 1071
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db GCAAGAGGACAGCTTATGAATCTTCAAGATAATTTGAATAAAGCAAGTATTGACAGC 1131
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db TATTGGAAGAGTGGGCACAAACCAGATAATATTAAAGGGAATTTGGAATTCGAATGTT 1191
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db CACTTCAGTTACCATCTCGAAAAGAAGTTAAGATCTTGAAGGCGCTGAACCTGAAGGTG 1251
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrVal 439
Db CAGAGTGGGACAGCGGTGGCCCTGGTTGGAAACAGTGTGGTGGGAAGAGCACACAGTCT 1311
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyLysAsp 459
Db CAGCTATGTCAGAGGCTCTATGACCCACAGAGGGATGGTCAGTGTGATGGACAGGAT 1371
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db ATTAGGACCATAAATGTAAGTTTCTACGGGAATCATTTGGTGTGATGGAGAACCT 1431
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db GTATTGTTTCCACACGATAGTGAACCAATTCGCTATGGCGGTGAAATGTCACCATG 1491
QY 500 AspGlnIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db GATGAGATTGAGAAAGCTGTCAAGGAGCCATTCCTATGACCTTTATCATGAACCTGCCT 1551

QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db CATAAATTTGACACCCCTGGTTGGAGAGAGAGGGGCCCGCTTGTAGTGTGGGACAGCAG 1611
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db AGGATCGGCATTGACGCTGCTGTTGCGAACCCCAAGATCTCCCTGCTGGATGAGGCC 1671
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db ACGTCAGCCTTGGACACAGAAAGCGAAGCAGTGTTCAGGTGGCTCTGGATAAGCCAGA 1731
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db AAAGTCGCGACCAACCATTTGTGATAGCTCATCGTTTGTCTACAGTTCGTAATGCTGACGTC 1791
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db ATCGCTGGTTTCGATGATGGAGTCATTGTGGAGAAAGAAATCATGATGAACCTCATGAAA 1851
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db GAGAAAGGCATTTACTTCAAACTTGTCAATGCGACAGCAGGAAATGAAGTTGAATTA 1911
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db GAAAATGCGAGCTGATGAATCCAAAAGTCAAAATGTGCTTGGAAATGCTCTCAAAATCAT 1971
QY 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db TCAAGATCCAGCTTAATAAGAAAAGATCAACTCGTAGGAGTGTCCGTGGATCAACAGCC 2031
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGTATACCTCCAGTTTCCTTT 2091
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db TGGAGATTATGAAGCTAAATTAACCTGAATGGCTTATTTTGTGTGTGTGTGTGTGTGTGT 2151
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db GCCATTATAAATGGAGGCTCGAACCGACATTTGCAATAATATTTTCAAGAGTATTAGGG 2211
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db GTTTTACAGAATTGATGATCCTGAAACAAACACAGACAGAAATAGTAATCTGTTTCACTA 2271
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db TTGTTTCTAGCCCTTGGAAATTTATTTCTTTTATACATTTTTCCTTCAAGGTTTCCACATT 2331
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db GGCAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATGTTTCCGATCCATGCTC 2391
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db AGACAGGATGTGAGTGTGTTGATGACCTAAAAACACCACTGGAGCATTTGACTACCAGG 2451
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db CTCGCCAATGATGTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTGTGCTGTAATTAACC 2511
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db CAGAAATAGCAAAATCTTGGGACAGGAATAATATATCTTCATCTATGTTGGCACTA 2571
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db ACACCTGTTACTCTTAGCAATTTGCCCATCATTTGCAATAGCAGGAGTGTGTGAAATGAAA 2631

QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
|||||
Db 2632 AFGTTGCTGGCAAGCACTGAAGATAGAAAGAACTAGAGAGTGTGGGAAGATCGCT 2691
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
|||||
Db 2692 ACTGAAGCAATAGAAAACCTCCGAACCGTGTCTTCTTGACTCAGGAGCAAGCTTGAA 2751
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2752 CATATGATGTCAGAGTTTCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATC 2811
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 2812 TTTGGAATACATTTCTCCACCAGGCAATGATATTTTCTATGCTGGATGTTTC 2871
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
|||||
Db 2872 CGGTTGGCGCTACTTGGTGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 2931
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 2932 TTTTCAGCTGTGCTTTGGTGCCATGGCGGTGGGCAAGTCAGTTTCATTTGCTCCCGAC 2991
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 2992 TATGCCAAGCCAAATATACAGCCACCATCATCATGATCATGTGAAAAACCCCTTG 3051
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3052 ATTGACAGCTACAGCAGCGGAGGCTAATGCCGAACACATTGGAAGGAAATGTCACATTT 3111
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3112 GGTGAAGTGTGATTCAACTATCCCAACCGCCGACATCCAGAGTTCAGGGGACTGAGC 3171
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3172 CTGGAGGTGAAGAGGGCCACAGCTGGCTCTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3231
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3232 ACAGTGGTCCAGCTCCTGGAGCGGTCTACGACCCCTTGGCAGGGAAGTGTCTGTGAT 3291
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3292 GGCAAGAAATAAAGCGACTGAATGTTTCAGTGGCTCCGAGCACACCTGGGCATCGTGCC 3351
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3352 CAGGAGCCATCCTGTTGACTGCGACATGCTGAGAACATGCTATGGAGACAACAGC 3411
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3412 CGGGTGGTGTACAGAAAGAGATCGTAGGGCAGCAAGAGGCGCAACATACATGCGCTC 3471
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3472 ATFCAGTCACTGCCCTTAATAATATATAGCACTAAAGTAGGAGACAAGGAACCTCAGCTCTCT 3531
QY 1180 GlyGlyLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3532 GTGGCCAGAAACAACGATGTCATAGCTGTGGCCCTGTTGTAGACGCCATATTTTG 3591
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3592 CTTTGGATGAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAAGCC 3651
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3652 CTGGCAAGCCAGAGAGGGCGCACCTGCATTTGTGATTGCTCAGCCGCTGTCCACCATC 3711
QY 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259

Db 3712 CAGAATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGAGTCAAGGAGCATGGCAGCAT 3771
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3772 CAGCAGCTGTGGCAGACAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 3831
QY 1280 LysArg 1281
Db 3832 AAGCGC 3837
RESULT 10
E02326
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
OS Homo sapiens
PN JP 1990100680-A/1
PD 12-APR-1990
PF 05-OCT-1988 JP 1988251475
PI UEDA KAZUMITSU, KOMANO TORU
PC C12N15/12,C12N1/21,C12Q1/68;
CC strandedness: Single;
CC topology: Linear;
CC *source: tissue_type=Adrenal gland;
CC *source: clone=SAM1132;
FH Key
FT 5'UTR
FT CDS
FT mat_peptide
FT
FT 3'UTR
source
location/Qualifiers
1..4378
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1315 a 818 c 1057 g 1188 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4378
Score: 5863.00 Matches: 1163
Percent Similarity: 95.09% Conservative: 56
Best Local Similarity: 90.72% Mismatches: 59
Query Match: 90.56% Indels: 4
DB: Gaps: 4
US-09-672-725C-23 (1-1281) x E02326 (1-4378)
QY 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 138 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAGAACTTTTTTAAACTGAAC 197
QY 20 LysLysSerLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 198 AATAAAGTGAAGAAA---GATAAGAGAGAAAGAAACCACTGTCTGATTTTCAATG 254
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
|||||

Db	255	TTTCGCCTATTCAAATTGGCTTGACAAAGTTGTATATAGTGGTGGGAACCTTTGGCTGCCATC	314
Qy	60	IleHisGlyAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe	79
Db	315	ATCCATGGGCTGGACTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT	374
Qy	80	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr	99
Db	375	GCAAAATCGAGGA--AATTTAGAAGATCTGATGTCACAAACATCAGCTAAATAGAAGTGAATC	431
Qy	100	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr	119
Db	432	AATGATCAGGGTCTCTATGAAAT---CTGGAGGAGACATGACACAGGTATGCCATATAT	488
Qy	120	TyrSerGlyIleGlyAlaGlyValLeuValAlaTyrIleGlnValSerPheTrpCys	139
Db	489	TACAGTGGAAATGGTCTGGGTGCTGTTGCTTACATTCAGTTTCATTTTGGTGC	548
Qy	140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159
Db	549	CTGCGAGCTGGAAGACAAATACACAAAATTAGAAACAGTTTTTTTCATGCTATAATGCGA	608
Qy	160	GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179
Db	609	CAGGAGATGGCTGGTTGATGTGCACGATGTGGGAGCTTAACACCCGACTTACAGAT	668
Qy	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle	199
Db	669	GATGCTCCAAGATTATGAGGAATTGGTGACAAAATTGGAAATGTTCTTCAGTCAATG	728
Qy	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219
Db	729	GCAACATTTTTTCACTGGTTTTATGATAGGATTTACACGTGGTGGAACTAAACCTTGTG	788
Qy	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239
Db	789	ATTTTGGCCATCAGTCTCTCTTGACATGTCAGCTGCTGCTGGGCAAGATACTACTCT	848
Qy	240	SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal	259
Db	849	TCATTTACTGATTAAGACTCTTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGCT	908
Qy	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr	279
Db	909	TTGGCAGCAATTAGAAGCTGTGATTGCTATTGGAGCACAAAAGAAAGACTTGAAGGTAC	968
Qy	280	AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299
Db	969	AACAAAAATTTAGAGAGAGCTAAAAGATTTGGGATAAAGAAAGCTATTACAGCCAAATT	1028
Qy	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319
Db	1029	TCATATAGGTGCTCTTCTCTGATCTATGATCTTTATGCTCTGGCCCTCTCGTATGGG	1088
Qy	320	ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer	339
Db	1089	ACCACCTTGTCCTCTCAGGGGAATATCTATTGGACAAGTACTCACTGTATCTTTTCT	1148
Qy	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359
Db	1149	GTATTAATTGGGCTTTTATGTTGGCAGGCAATCTCCAAGCAATGTAAGCAATTTTGCAAT	1208
Qy	360	AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379
Db	1209	GCAAGAGGACAGCTTATGAAATCTTCAAGATTAATGATAATAGCCAAAGTATTGGACAGC	1268
Qy	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399
Db	1269	TATTCGAAGAGTGGGCACAAACAGATAATATAAGGGAAATTTGGRAATTCAGAAATGTT	1328
Qy	400	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419
Db	1329	CACCTCAGTTACCCTCTCGAAAGAAGTTAAGATCTTTGAAGGGTCTGAACCTGAAGGCT	1388

Qy	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal	439
Db	1389	CAGAGTGGCAGACGGTGGCCCTGGTTGGAAACACAGTGGCTGTGGGAAGACAGACACAGTC	1448
Qy	440	GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459
Db	1449	CAGCTGATGTCAGAGGCTCTATGACCCACAGAGGGATGGTCAGTGTGTGATGGACAGGAT	1508
Qy	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro	479
Db	1509	ATTAGGACCATAAATGTAGGTTTCTACGGGAAATCATTTGGTGTGGTCAGTCAGGAACCT	1568
Qy	480	ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499
Db	1569	GTATGTTTGGCCACACAGTAGCTGAAACATTCGCTATGGCCGTGAAATGTCAACCATG	1628
Qy	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro	519
Db	1629	GATGAGATTGAGAAGCTGTCAAGGAAGCAATGCCCTATGACTTTATCATGAACATGGCT	1688
Qy	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539
Db	1689	CATAAATTGACACCCCTGGTTGGAGAGAGAGGGGCCAGTTGAGTGGTGGCGACGAAGCAG	1748
Qy	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559
Db	1749	AGGATGCCCATTCGACGTGCCCTGGTTGCTGTAACCCACAGATCCTCCTGCTGGATGAGGCC	1808
Qy	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579
Db	1809	ACGTGACGCTTGGACACAGAAACCGAAGCAGTGGTTGAGTGGCTCTGATGAAGGCCAGA	1868
Qy	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599
Db	1869	AAAGGTGGGACACCACTTGATGATGCTCATCTGTTGCTTACAGTTGCTGAATGCTGACGTC	1928
Qy	600	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619
Db	1929	ATCCCTGGTTTCGATGATGAGTCACTGTGTGGAAAGGAATCATGATGAATCATGAA	1988
Qy	620	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu	639
Db	1989	GAGAAAGGCATTTACTTCAACCTGTGCACAAATGCACAGACAGAGAAATGAAGTTGAATTA	2048
Qy	640	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	659
Db	2049	GAANAATGCAGCTGATGATCCAAAAGTGAATTTGATGCTTGGAAATGCTCTCAAAATGAT	2108
Qy	660	SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly	679
Db	2109	TCAAGATCCAGCTAATAAGAAAAGATCAACTCGTAGGAGTGTCCGTGGATACAAAGCC	2168
Qy	680	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	699
Db	2169	CAAGACAGAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCTCTTT	2228
Qy	700	TrpArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCys	719
Db	2229	TGGAGGATTATGAAGCTAAATTTAACTGAATGSCCTTATTTTGTGTTGGTGATATTTGT	2288
Qy	720	AlaIleIleAsnGlyLysLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	739
Db	2289	GCCATTATAATGGAGCCTGCACACAGATTTGCAATAATATTTTCAAGATTTATAGGG	2348
Qy	740	IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal	759
Db	2349	GTTTTTACAGAAATTGATGATCCTGAAACAAACACACAGATAGTAGTACTGTTGTTCACTA	2408
Qy	760	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	779
Db	2409	TGTGTTCTACGCTTGGAAATATTTCTTTTATATATATATTTCTTTCAGGGTTCACATTT	2468

QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
|||||
Db 2469 GGCAAGCTGGAGAGATCCTCACCAGCGCTCCGATACATAGTGTTCCTCCATCCATGCTC 2528
QY 800 ArgGlnAspValSerTrpPheAspSerProLysAsnThrThrGlyAlaLeuThrThrArg 819
|||||
Db 2529 AGACAGGATGTGAGTGGTTGATGACCTAAACACACACCTGGAGCATGTGACTACCAAGG 2588
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
|||||
Db 2589 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCTGTATTACC 2648
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
|||||
Db 2649 CAGAAATATAGCAAACTCTGGGACAGAAATATATATCTTCATCTATGTTGGCAACTA 2708
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
|||||
Db 2709 ACACCTGTTACTCTTAGCAATGTGACCAATGTCATTCATGTCATGAGGAGTGTGGAATGAA 2768
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
|||||
Db 2769 ATGTTCTGGACAAGCACTGAAAGATAAGAAAGAACTAGAAGTTCCTGGGAAGATCGCT 2828
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
|||||
Db 2829 ACTGAAGCAATAGAAAACCTCCGAACCGTGTCTTCTGACTCAGGACAGAAAGTTTGAA 2888
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
:::|||||
Db 2889 CATATGTATGCTCAGAGTTGACAGGTACCATACAGAAACTCTTTGAGGAAGACACATC 2948
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 2949 TTTTGAATTACATTTCTTCACCCAGGCAATGATGATTTTCTCTATGCTGATGTTTC 3008
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
|||||
Db 3009 CGGTTTGGACCTACTGTGGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 3068
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
|||||
Db 3069 TTTTCAGCTGTGCTTTGTGGTCATGCGCGTGGGCAAGTCAATTCATTTGCTCTGAC 3128
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleLeuLysSerProLeu 1019
|||||
Db 3129 TATGCCAAAGCCAAATATCAGCAGCCCATCATCATCATGATCATTTGAAAAACCCCTTG 3188
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
|||||
Db 3189 ATTGACAGCTACAGCAGGGAAGGCTTAATCCGAACACATGGAAGAAATGTCATTT 3248
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
|||||
Db 3249 GGTGAAGTTGATTCAACTATCCACCCGACCGGACATCCAGTCTTCAGGACCTGAGC 3308
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
|||||
Db 3309 CTGGAGGTGAAGAGGCCACAGCTGGCTGTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3368
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
|||||
Db 3369 ACAGTGTGTCAGCTCTGGAGCGGTTCTACGCCCTTGGCAGGGAAGTGCCTGTGAT 3428
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
|||||
Db 3429 GGCAAGAAATAAGCCGACTGAATGTTTCAGTGGCTCCGAGCACACCTGGCATCGTCTCC 3488
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
|||||
Db 3489 CAGGAGCCCATCTGTTGACTGACGATTTGCTGAGAACATTTGCTATGGAGACCAAGC 3548
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159

Db 3549 CGGTGTGTGTCCAGGAAGAGATTGTGAGGCGACCAAGGAGGCCACATACATGCTTC 3608
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
|||||
Db 3609 ATCAGTCACTGCCTAATAATATAGCACATAAGTAGGAGACAAAGAACTAGCTCTCT 3668
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
|||||
Db 3669 GGTGGCCAGAAACACCATTCCTAGCTGCTGCCCTTGTTCAGACGCTCATATTTTG 3728
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
|||||
Db 3729 CTTTGGATGAAGCCACGCTCAGCTCGATACAGAAAGTGAAGTTGTCACCAAGGCC 3788
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 3789 CTGGACAAAGCCAGAGAGCGCCCTGCTGATGCTGCTCACCCTGCTCCACCATC 3848
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||
Db 3849 CAGATGACAGACTTATATAGTGTGTTTCAGATGCGCAGAGTCAAGGACATGGCACCAT 3908
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 3909 CAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGTCAGTGTCCAGCTGGAACA 3968
QY 1280 LysArg 1281
|||||
Db 3969 AAGCGC 3974
RESULT 11
AXI08654 AXI08654 4186 bp DNA linear PAT 30-APR-2001
LOCUS Sequence 1 from Patent WO0123565.
DEFINITION AXI08654
ACCESSION AXI08654
VERSION AXI08654.1 GI:13923886
KEYWORDS
SOURCE crab-eating macaque.
ORGANISM Macaca fascicularis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (bases 1 to 4186)
Stocker, P.J., Steimel-Crespi, D.T. and Crespi, C.L.
P-glycoproteins from macaca fascicularis and uses thereof
Patent: WO 0123565-A 1 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
Location/Qualifiers
1..4186
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
100..3942
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37764.1"
/translation="MDLEGRNGAEKNKFKLNKSKKKKPKTPVSVFSPFRYSN
WLDKLYWVGTLLAIIHGAGIPLMWLVGDMTDTFANAGNLGDLGALLTNSNITDTV
GWFVDHVGELNTRTDDVSKINIGIKGIFQFFQSMATFFTFIVGFTGKWKLTLYI
LAISPLVGLSAAVWAKILSFTDKELAYAKAGAAVEELAAIRTVIAFGQKKELER
YKNLEAKRIGIKKATITANISGAFLIYASVALAFYWTLLVLSKEYSGQGLVY
PFSVLIGAFSVGQASPSIEAFANARGAAFEIKIDNKPISDSYKSKGKPDNKGKL
BFRNVHFSYPSKKEVKILGKLVQSGQTVKLVNKGSCGKSTVQVMORLYDPTGEM
BYVDGDIRTINVRLEIIGVVSQEPVLPAATTAENIRYGRDVTMDIEISKAVKEAN
AYDFLTKLPQKRTTIVIAHRLSTVRNADYIAGFDGVIKEKHDELMEKGIYFK
AVVQVALDKARKRTTIVIAHRLSTVRNADYIAGFDGVIKEKHDELMEKGIYFK
LVTMOTAGNELENADESEIDLEMSHSDSSSLIRKSTRSRVSRSQGDRLK
SPTKALDESIPPSFWPRIMKLNLTWPFVVGVAINGLQIPAFVAFVSKIGIFT
RNDDETQRQSNLFLSLFLVIGVSTFTFQGTGKAGEILTKRLRYWVFRMLR
QDVSFDDPKNTTGALTITRLANDAAQVKAGISRLAITITONIANGLGIIISLYGWO
ITLTLALLAIVPIIATAGVVMKMLSGQALKDKKELEGAKIATEALENFRVTVSLTQEO

KFEHMDQSLQVYRNSLRKAHIFGIFTSFTQAMFYSVAGCFRFGAYLVAHLSMFE
DVLVFSAVVFGAMVQSVFAPDYAKAKVSAHHIMIEKPLIDSYSTEGLEKENT
LEGNTFNEVVFVYTRLDIPVLQGLSLEVKKGOTLALVSSCGSKSTVVOLLREYD
PLAGKVLDDKEIKQLNVOVLRHGLTVSOEPLIFDCSISENIATYGDNSRVYSOEIV
RAKAEANIHFISLPNKYSTRVGDGTQLSGQKQRIALARALVQPHILLIDERIS
ALDTESKVVQVEALDKAREGRTCIIVIAHRLUSTIONADLIIVFONGRVKBEHGHQULLA
QKGIYFSWVSQAGAKRQ"

BASE COUNT 1226 a 799 c 1039 g 1122 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4186
Score: 5862.00 Matches: 1163
Percent Similarity: 95.32% Conservative: 59
Best Local Similarity: 90.72% Mismatches: 56
Query Match: 90.55% Indels: 4
DB: 6 Gaps: 4

US-09-672-725c-23 (1-1281) x AX108654 (1-4186)

QY 1 MetAspProGluGlyArgLysGlySerAlaGlu---LysAsnPheTrpLysMetGly 19
Db 100 ATGGATCTTGAAGGCGCCGATGGAGGACGAGAGAGAACITTTTAACTGAAC 159
QY 20 LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 160 AATAAAGTAAAAA---GATAAGAGAGAAAGAAACCAACTGTCAGTGTATTTCAATG 216
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 217 TTTCGCTATTCAATTTGGCTTGACAGTGTATGTGTGGTGGAACTTTGGCTGCCATC 276
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 277 ATCCATGGAGCTGGACTCTCTCATGATGCTGGTGTGGAGACATGACGGATACCTTT 336
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 337 GCNAATGCAGGA---AATTAGGAGATTTAGGAGCTCTGTGTGACTAATAGCAGTAATATC 393
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 119
Db 394 ACTGATACAGTGCCTGGTGCATGAAT---CTGGAGAGAGATATGACAGGTATGCCATTATAT 450
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 451 TACAGTGGAAATTTGGTGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 510
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 511 CTGGCAGCTGGAGAGACAATACACAAATTTAGAAAACAGTTTTTTCATGCTATAATGCCA 570
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 571 CAGGAGATAGCTGGTTTGATGTGCACGATGTGGGAGCTTAAACCCGGCTTACAGT 630
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db 631 GATGTCTCCAAAGATTAAATGAAGGAATTTGGTGACAAATTTGGAATGTCTTTTCAGTCAATG 690
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 691 GCACATTTTTCCTGGGTTTATAGTAGGATTTTACACGTGGTGGTGGTGGTGGTGGTGGTGG 750
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 751 ATTTTGGCCATAGTCTCTTCTTGGAGCTGTCAGCTGCGAGTCTGGGCAAGATACATGCTCT 810
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 811 TCATTTTACTGTATAAAGAACTCTTAGCTTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 870
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysLysLysLysLysLys 279

Db 871 TTGGCAGCAATTAGAACTGTGATTGCTTTGGAGGACAAAAGAAAGCAACTCCAAAGGTAC 930
QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 931 AACAAAAATTTGAAGAAGCTAAAGAATTTGGGATAAAGAAAGCTATTACAGCAATATT 990
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 991 TCTATAGGTGCTGCTTCTCTGCTTATGCTATGCTATGCTGCTGCTGCTGCTGCTGCTGCTG 1050
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 1051 ACCACCTTGGTCTCTCAAGGAATATCTATTGGACAAGTACTCCTGCTATCTCTTCT 1110
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1111 GTATTAATTTGGGCTTTTGTAGTTGGACAGGCATCTCCAAGCATTTGAAGCATTTGCAAT 1170
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1171 GCAAGAGAGGAGCTTTTGAATCTTCAAGATAATTAAGGAAATTTGGAATTCAGAAATGTT 1230
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1231 TATTCGAAGAGTGGGCACAAACACAGATAATTAAGGAAATTTGGAATTCAGAAATGTT 1290
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1291 CACTTTCAGTTACCCATCTCGAAAAGAGTTAAGATCTTTGAAGGGCTTGAAGCTGAAGGTG 1350
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1351 CAGAGTGGCAGACGGTGGCCCTGTTTGGAAACACGGCTGTGGGAAGAGCACAAACGGTC 1410
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1411 CAGCTGTATGCAGAGCTTTATGACCCACAGAGGCGATGGTCAGTGTGATGGACAGGAT 1470
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1471 ATTAGACCATTAACGTAAAGTTTCTACGGAAATCATCGGTGGTGGTGGTGGTGGTGGTGG 1530
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1531 GTATTTGTTGCCACACCATAGCTGAAACATTCGCTATGCTGCTGCTGCTGCTGCTGCTGCT 1590
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1591 GATGAGATTGAGAAGCTGTCAAGGAAGCCAAATGCCCTATGACTTTATCATCAAACTGCCT 1650
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1651 CAGAAATTTGACACCTGTTGGAGAGAGAGGGGCCAGCTGAGTGGTGGTGGTGGTGGTGG 1710
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAla 559
Db 1711 AGGATCGCAATTCAGCTGGCTGCTGTTGCGAACCCCAAGATCTCTCTGCTGCTGCTGCTGCT 1770
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 1771 ACCTGACCTTTGGACACAGAAAGTGAAGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1830
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 1831 AAAGGTGGACCCACCATTTGATAGCTCATCTGTTGCTAGGGTTCGTAATGCCGAGCTC 1890
QY 600 IleAlaGlyPheAspAspGlyValIleValLysGlyAsnHisAspGluLeuMetLys 619
Db 1891 ATCGCTGTTTCGATGATGGAGTCATTGTGAGAAAGAAATCATGATGATGATGATGATGATG 1950
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639

QY	113	MetThrThrTyrAlaTyrTyrTyrSerGlyIleGlyValLeuValAlaAlaTyr	133
DB	439	ATGACCAAGGTATGCTATTATTACAGTGGAAATTTGGTGCTGGGTGCTGCTGCTGCTTAC	498
QY	133	IleGlnValSerPheTrpCysLeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGln	152
DB	499	ATTCAAGTTCATTTTGGTCCTGGCAGCTGGAAGACAAATACACAAAATTAGAAAACAG	558
QY	153	PhePheHisAlaIleMetArgGlnGluIleGlyTrpPheAspValHisAspValGlyGlu	172
DB	559	TTTTTTTCATGCTATAATGACACAGAGATAGGCTGGTTGATGTGCACGATGCTGGGAG	618
QY	173	LeuAsnThrArgLeuThrAspAspValSerLysIleAsnGluGlyIleGlyAspLysIle	192
DB	619	CTTAACACCCGGCTTACAGATGATGCTCCAAGATTAAAGAGGAATTTGGTGACAAAATT	678
QY	193	GlyMetPheGlnSerIleAlaThrPhePheThrGlyPheIleValGlyPheThrArg	212
DB	679	GGAAATGTTCTTCAGTCAATGGCAACATTTTTCACGTGGGTTTTATAGTAGGATTTACAGT	738
QY	213	GlyTrpLysLeuThrLeuValIleLeuAlaIleSerProValLeuGlyLeuSerAlaAla	232
DB	739	GGTTGGGAAGCTAACCCCTTGATTTTGGGCATCAGTCCCTGTTTCTTGGACTGTGCACGTCA	798
QY	233	IleTrpAlaLysIleLeuSerSerPheThrAspLysGluLeuLeuAlaTyrAlaLysAla	252
DB	799	GTCGTGGCCAAAGATACGTCTTCATTTACTGATAAAGAACTCTTACGTATATGCAAAAGCT	858
QY	253	GlyAlaValAlaGluGluValLeuAlaAlaIleArgThrValIleAlaPheGlyGlyGln	272
DB	859	GGAGCAGTAGCTGAAGAGGTCTTGGCAGCAATTAGAACTGTGATTTGCATTTGGAGGACAA	918
QY	273	LysLysGluLeuGluArgTyrAsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLys	292
DB	919	AAGAAAGAACTCGAAAGGTACAAACAAAATTTAGAAAGAACTTAAAGAAATTTGGGATAAG	978
QY	293	LysAlaIleThrAlaAsnIleSerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyr	312
DB	979	AAAGCTATTACAGCCAATATTTCTATAGGTGCTGCTTCTCTCTATCTATCTATGTCATCTAT	1038
QY	313	AlaLeuAlaPheTrpTyrGlyThrSerLeuValLeuSerSerGluTyrSerIleGlyGln	332
DB	1039	GCTCTGGCCCTCTGGTATGGACCACTTGGTCTCTCAAGGAATATTCTATTGGACAA	1098
QY	333	ValLeuThrValPhePheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerPro	352
DB	1099	GTACTCACTGATCTCTTTCTGTATTAAATGGGGCTTTTAGTGTGGACAGCATCTCCA	1158
QY	353	SerIleGluAlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheLysIleIleAsp	372
DB	1159	AGCATTAAGCAATTTGCAATGCAAGAGGAGCAGCTTTTGAAATCTTCAAGATATTGAT	1218
QY	373	AsnLysProSerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGly	392
DB	1219	AATAAGCCAAGTATTACAGCTATTTCGAAGAGTGGGCACAAACACAGATAAATATTAAGGA	1278
QY	393	AsnLeuGluPheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeu	412
DB	1279	AAATTTGGAATTCAGAATGTTCACTTCAGTTTACCATCTCTGAAAGAAGATTTAAGATCTTG	1338
QY	413	LysGlyLeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGly	432
DB	1339	AAGGGCTGAACTGAAGGTGCAGAGTGGCAGACGGTGGCCCTGGTTTGGAAACACGGCGC	1398
QY	433	CysGlyLysSerThrThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMet	452
DB	1399	TGTGGGAAGACACACAGCTCCAGCTGATGCAGAGGCTTTATGACCCACACAGGGCATG	1458
QY	453	ValCysIleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThr	472
DB	1459	GTCAGTGTGATGGACAGGATATTAGGCACATAAAGCTTAAGGTTTCTACGGGAATCATC	1518

Qy 473 GlyValValSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyr 492
Db 1519 GGTGGTGGTGGAGAACCTGTATTGTTGCCACCGATAGCTGAAACATTTCGGTAT 1578
Qy 493 GlyArgGluAsnValThrMetAspGluIleGluLysAlaValLysGluAlaAsnAlaTyr 512
Db 1579 GGTGCTGAAGATGTCACCATGGATGAGATTGAGAAAGCTGTCGAAGGAAGCCAATGCGCTAT 1638
Qy 513 AspPheIleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGln 532
Db 1639 GACTTTATCATGAACATGCTCAGAAATTCACACCTGTTGGAGAGAGGGGCCAG 1698
Qy 533 LeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLys 552
Db 1699 CTGAGTGGTGGCAGAGACAGAGATCGCCATTGACAGTGCCTGGTTCCACACCCCAAG 1758
Qy 553 IleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGln 572
Db 1759 ATCTCTCTGTCGACGAGGCCACGTCAGCCTTGGACACAGAAAGTGAAGCAGTGGTTTCAG 1818
Qy 573 ValAlaLeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSer 592
Db 1819 GTGGCTCTGATAGGCCAGAAAGTCCGACCAACCATTTGATAGCTCATCTGTTGTCT 1878
Qy 593 ThrValArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGluLysGly 612
Db 1879 ACGTTTCGTAAATGCCGACGTCATCGCTGGTTTCGATGATGGAGTCATTGTGGAGAAAGGA 1938
Qy 613 AsnHisAspGluLeuMetLysGluLysGlyIleIleIleIleIleValThrMetGlnThr 632
Db 1939 AATCATGATGAGCTCATGAAGAGAAAGGCAATTTACTTCAAACTGTGCACAATGCAGACA 1998
Qy 633 ArgGlyAsnGluIleGluLeuGluAsnAlaThrGlyGluSerLysSerGluSerAspAla 652
Db 1999 GCAGGAATGAATGTGAATAGAAATGCAGCTGATGAATCCAAAGTGAATTTGATACC 2058
Qy 653 LeuGluMetSerProLysAspSerGlySerLeuIleLysArgArgSerThrArgArg 672
Db 2059 TTGGAATGCTTCACATGATTACGAGTCCAGTCTATAGAAAAGATCCATCGTAGG 2118
Qy 673 SerIleHisAlaProGlnGlnAspArgLysLeuGlyThrLysGluAspLeuAsnGlu 692
Db 2119 AGTGTGCTGGATCACAAGGCAGACAGAAAGCTTAGTACCAAGAGAGCTCTGGATGAA 2178
Qy 693 AsnValProValSerPheTrpArgIleLeuLysLeuAsnSerThrGluTrpProTyr 712
Db 2179 AGTATCTCCAGTTCTCTTTGGAGGATTAAAGCTAAATTTAACTGAGTGGCCTTAT 2238
Qy 713 PheValValGlyIlePheCysAlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIle 732
Db 2239 TTTGTTGTTGGTGTATTTTGTGCCATTATAAATGGAGCTCTGCAACAGCATTTGCAGTA 2298
Qy 733 IlePheSerArgIleIleGlyIlePheThrArgAspGluAspProGluThrLysArgGln 752
Db 2299 ATATTTTCAAGATTATAGGATTTTACAAGAAATGATGATGCCGAACAAAAACGACAG 2358
Qy 753 AsnSerAsnMetPheSerValLeuPheLeuValLeuGlyIleIleSerPheIleThrPhe 772
Db 2359 AATAGTAACCTGTTGTTTCTACATTGTTCTAGTCTCTTGGAAATTTGTTCTTTTATACATTT 2418
Qy 773 PheLeuGlnGlyPheThrPheGlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyr 792
Db 2419 TTCCTTCAGGCTTTCACATTTGGCAAGCTGGAGAGATCTCACCAGCGCTCCGATAC 2478
Qy 793 MetValPheArgSerMetLeuArgGlnAspValSerTrpPheAspAspProLysAsnThr 812
Db 2479 ATGCTTTTCCGATCCATGCTCAGACAGATGTGAGCTGGTTTGATGACCCCTAAAAACACC 2538
Qy 813 ThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaGlnValLysGlyAlaIleGly 832
Db 2539 ACTGAGCATTTGACTACAGGCTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGT 2598
Qy 833 SerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSer 852

Db 2599 TCCAGGCTTGCATATATTACCAGATATAGCAATCTTGGACAGGAATATATTATATCC 2658
Qy 853 LeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAlaIleValProIleIleAlaIle 872
Db 2659 TTAATCTATGTTGGCAACTGACACTGTACTTCTTAGCAATTTGTACCCATCATTCGAATA 2718
Qy 873 AlaGlyValValGluMetLysMetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeu 892
Db 2719 GCAGGAGTGTGTAATGAAATGTTGCTGGACAGCACTGAAAGATAAAGAAAGACTA 2778
Qy 893 GluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeu 912
Db 2779 GAAGTGTGGGAAGATCCTACTGAAGCAATAGAAAACCTCCGAACCTGTTGTTCTTTG 2838
Qy 913 ThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsn 932
Db 2839 ACTCAGGACGAAGTTTGAACATATGATGATCAGAGTTTTCAGGTACCATACAGAACA 2898
Qy 933 SerLeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyr 952
Db 2899 TCTTTGAGGAAGCACACATCTTTGGAATCACGTTTTCTTCCAGCAGCAATGATGAT 2958
Qy 953 PheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsn 972
Db 2959 TTTTCTATGCTGGATGTTTCGGTTTGAGCCTACTTGTGCGACATAGTCTCATGAGC 3018
Qy 973 PheGlnAspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAlaValGln 992
Db 3019 TTTGAGGATGTTCTGTAGTATTTTTCAGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3078
Qy 993 ValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMet 1012
Db 3079 GTCAGTTTCATTTGCTCTGCTGATGCCAAGCCCAAGTATACAGCCGCCACATCATCATG 3138
Qy 1013 IleIleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThr 1032
Db 3139 ATCATTTGAAAAACCCCTTTGATTGACACTACAGCACAGAGGCCCTAAAGCCGAACACA 3198
Qy 1033 LeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIle 1052
Db 3199 TTGGAAGGAAATGTCACATTTAATGAAGTTGATTCAACTATCCACCCGACTGGACATC 3258
Qy 1053 ProValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrLeuAlaLeuValGly 1072
Db 3259 CCAGTCTTCCAGGGCTGAGCCTGGAAGTGAAGAGGGCCAGCCGCTGCCCTGTGGGC 3318
Qy 1073 SerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeu 1092
Db 3319 AGCAGTGGCTGTGGGAAGACACGGTGGTCCAGCTCTGAGCGGTTCTATGACCCCTTG 3378
Qy 1093 AlaGlySerValLeuLeuAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArg 1112
Db 3379 GCGGGGAAAGTGTGCTTACGCGCAAGAAATAAAGCAACTGAATGTTTCAGTGGCTCGGA 3438
Qy 1113 AlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsn 1132
Db 3439 GCACACTGGGATCGTGTCCAGGAGGCCATCTCTGTTGACTGCGCATGATGAGAAC 3498
Qy 1133 IleAlaTyrGlyAspAsnSerArgValValSerHisGluGluIleMetGlnAlaAlaLys 1152
Db 3499 ATTGCTATGGAGACACACCGGGTGGTGTTCACAGGAAGAGATCGTGAGGCCACCCAAG 3558
Qy 1153 GluAlaAsnIleHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGly 1172
Db 3559 GAGGCCAATATACAGCCCTTCATCGAGTCAGTGCCTTAATAAATATAGCACCAGAGTAGGA 3618
Qy 1173 AspLysGlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeu 1192
Db 3619 GACAAAGGAACACAGCTCTCTGTGGTGGCCAGAAACACGCAATGTCATAGCTGTCGCCCTT 3678
Qy 1193 ValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer 1212

Db 3679 GTTAGACAGCCCTCAATATTTTCTTTGGATGAAGCCACATCAGCTCTGGATACAGAAAGT 3738
Qy 1213 GluLysValValGlnGluAlaLeuAspLysAlaArgGluGlyAlaGlyThrCysIleValIle 1232
Db 3739 GAAAGGTTGTCCAGAGAGCCCTGGACAAAGCCAGAGAGGCCGTACCTGCTGATTTG 3798
Qy 1233 AlaHisArgLeuSerThrIleGlnAsnAlaAspLeuIleValValPheGlnAsnGlyLys 1252
Db 3799 GCTCACCAGCTGCTCCACCATCCAGAAATGCAGATTAATAGTGGTGTTCAGAAATGGCAG 3858
Qy 1253 ValLysGluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleThrPheSerMet 1272
Db 3859 GTCAAGGAGCAGCGCACACATCAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATG 3918
Qy 1273 ValSerValGlnAlaGlyAlaLysArg 1281
Db 3919 GTCAAGTGTCCAGGCTGGAGCAAAAGCCG 3945
RESULT 13
AX322789 3860 bp DNA linear PAT 07-JAN-2002
LOCUS AX322789 Sequence 3 from Patent WO0192877.
ACCESSION AX322789
VERSION AX322789.1 GI:18093767
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Sorrentino,B. and Schuetz,J.
TITLE Method of identifying and/or isolating stem cells
JOURNAL Patent: WO 0192877-A 3 06-DEC-2001.
ST. JUDE CHILDREN'S RESEARCH HOSPITAL (US)
FEATURES
source
1. .3860
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1135 a 746 c 956 g 1023 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3860
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.47% Indels: 4
DB: 6 Gaps: 4
US-09-672-725C-23 (1-1281) x AX322789 (1-3860)
Qy 1 MetAspProGluGlyGlyValArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1 ATGGATCTTGAAGGAGCCGCAATCGAGGAGCAAGAGAGAACTTTTAAACTGAC 60
Qy 20 LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 61 AATAAAGTGAAGAA---GATAAGAGAAAGAAACCAACTGTCAGTGTATTTTCAATG 117
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 118 TTTTCCTATTCAAATTTGGCTTGACAAAGTTGATATGGTGGGAACCTTTGGCTGCCATC 177
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 178 ATCCATGGGCTGGACTTCCCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 237
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 238 GCAAAATGCAGGA---AATTTAGAAGATCTGATGTCAAAACATCACTAATAAGAGTGATATC 294
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 119

Db 295 AATGATACAGGGTTCTTCATGAAT---CTGGAGGAAGACATGACCAGATATATGCTATTAT 351
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 352 TACAGTGAATTTGGTGGGTGCTGGTGTGCTGTACATTCAGGTTTCATTTTGGTGC 411
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 412 CTGCAGCTGGAAGACAAATACACAAATTAGAAAACAGTTTTTTTCATGCTATAATGCGA 471
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 472 CAGAGATAGGCTGGTGTGATGTGCACGATGTTGGGAGCTTAAACACCCGACTACAGAT 531
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 532 GATGCTCTTAAGATTAAATGAAGTTATTTGGTGACAAATTTGGAATGTTCTTCAGTCAATG 591
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 592 GCAACATTTTTCACCTGGGTTTATAGTAGGATTTTACACGTGGTGGAGCTAAACCTTGTG 651
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 652 ATTTGGCCATCAGTCTCTTGGACTGTGAGCTGTCTGGGCAAGATACATCTATCT 711
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluIleVal 259
Db 712 TCATTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 771
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysLysGluLeuArgTyr 279
Db 772 TTGCAGCAATTAGAAGCTGTGATTGCAATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 831
Qy 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 832 ACAAAAATTTTAGAAGAGCTAAAGAAATTTGGGATTAAGAAAGCTATTACAGCCAATAT 891
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 892 TCTATAGTCTGCTTCTCTGCTGATCTATGATCATCTATGCTCTGCTCTGCTATGGG 951
Qy 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 952 ACCACCTTGGTCTCTCAGGGGAATATCTTATTGGCAAGTACTACTGTATTCTTTCT 1011
Qy 340 ValIleIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1012 GTATTAAATGGGCTTTTAGTGTGGACAGGCATCTCCAAAGCATTTGAAGCATTTGCAAT 1071
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1072 GCAAGAGGAGCAGCTTATGAAATCTTCAAGATAATTTGATAATAAGCCAAGTATTGACAG 1131
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1132 TATTGGAAGAGTGGGCACAAACCAAGATAATATTAAAGGAAATTTGGAATTCAGAAATGT 1191
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1192 CACTTCAGTTACCCATCTCGAAAAGAAAGTTAAGATCTTGAAGGCCCTGAACCTGAAGGTG 1251
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1252 CAGAGTGGGCAGAGCGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGAGCACACAGTC 1311
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1312 CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGATGGTGTGTTGAGACAGGAT 1371
Qy 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1372 ATTAGGACCATAAATGTAAGTTTCTACGGGAATCATTTGGTGTGCTGAGTCAGGAACCT 1431

QY	480	ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499
DB	1432	GTATTGTTGGCCACACGATAGCTGGAACAATTGCGTATGCGCCGTGAATAATGTCACCATG	1491
QY	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro	519
DB	1492	GATGAGATTGAAAGCTGTCAAGGAAGCCAATGCGCTATGACCTTTATCATGAACATGCGCT	1551
QY	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539
DB	1552	CATAAATTTGACACCTGGTTGGAGAGAGAGGGGCCAGTTGAGTTGGTGGCGAAGAAGCAG	1611
QY	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559
DB	1612	AGGATCGCCATTGACGTGCCCTGGTTCCGAACCCCAAGATCCTCTGCTGGATGAGGCC	1671
QY	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579
DB	1672	ACGTCAGCCTTGGACACAGAAAGCAAGCACTGGCTTCAGGTGGCTCGGATAAGGCCAGA	1731
QY	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599
DB	1732	AAAGGTGCGACCACCATTTGTGATAGCTCATCTGTTTGTCTACAGTTCTGTAATGCTGAGCTC	1791
QY	600	IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619
DB	1792	ATCGCTGGTTTCGATGATGGAGTCAATCTCGAGAAGAAGAAATCATGATGAACTCATGAAA	1851
QY	620	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu	639
DB	1852	GAGAAAGGCAATTTACTTCAAACTGTCACATGACAGACAGAGAAATGAAGTTGAATTA	1911
QY	640	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	659
DB	1912	GAAATGCGACGTGATGAATCCAAAAGTCAAAATTGATGCCTTGGAAATGCTTCAAATGAT	1971
QY	660	SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly	679
DB	1972	TCAAGATCCAGTCTAATAGAAAAAGATCAACTCGTAGGAGTGCCTGGATGATCACAAGC	2031
QY	680	GlnAspArgLysLeuGlyThrLysGluAspLysAsnGlnValProProValSerPhe	699
DB	2032	CAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT	2091
QY	700	TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	719
DB	2092	TGGAGGATTATGAAGCTAAATTTAACTGAATGGCCCTTATTTGTTGGTGTGTTATTTGT	2151
QY	720	AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	739
DB	2152	GCCATTATAATGGAGCCCTGCAACGAGANTTTGCAATATATTTTCAAGATTTATAGGG	2211
QY	740	IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetSerVal	759
DB	2212	GTITTTTACAGAATTGATGATCCTGAAACAAACACACAGAAATAGTAGTTCGTTCACTA	2271
QY	760	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	779
DB	2272	TTGTTTCTAGCCCTTGGAAATTAATTTCTTTATTTACATTTTTCCTTCAAGGTTTTCACATTT	2331
QY	780	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	799
DB	2332	GGCAAAAGCTGGAGAGATCTCCACCAAGCGCTCCGATACATAGTTTCCGATCATCGCTC	2391
QY	800	ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg	819
DB	2392	AGACAGGATCTGAGTTGGTTTGATGACCCTAAAAACACCACCTGGAGCATTCGATACCAAGG	2451
QY	820	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	839
DB	2452	CTGCCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTTCAGGCTTCAGGCTTCAATATACC	2511

QY	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrglyTrpGlnLeu	859
DB	2512	CAGAATAAGCAAAATCTTGGACAGGAATAATATATCCTTCATCTATGGTTGGCAACTA	2571
QY	860	ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGlnMetIys	879
DB	2572	ACACITGTTACTCTTAGCAATTTGTACCCATCATTGCAATAGCAGGAGTTGTTGAAATGAAA	2631
QY	880	MetLeuSerGlyGlnAlaLeuIysAspIysGluLeuLeuGlyAlaGlyIysIleAla	899
DB	2632	ATGTTGTTCTGGCAAGCACTGAAGATAGAAAGAACTAGAAAGTGTCTGGGAAGATCGCT	2691
QY	900	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnIysPheGlu	919
DB	2692	ACTGAAGCAATAGAAACTTCGGAACCGTGTTCCTTTGACTCAGGACGACGAAGTTTGAA	2751
QY	920	TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgIysAlaHisIle	939
DB	2752	CATATGTATGCTCAGAGTTTGAGGTACCATACAGAAACTCTTTTGGAGAAAGCACACATC	2811
QY	940	PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe	959
DB	2812	TTTGGAAATTACAAATTTCTTCCACCAGGCAGATGATGATATTTTCTATGCTGGATGTTTC	2871
QY	960	ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheClnAspValLeuLeuVal	979
DB	2872	CGGTTTGGAGCCTACTTGTGTGCACATAAACTCATGAGCTTTGGAGATCTTCTGTAGTA	2931
QY	980	PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp	999
DB	2932	TTTTTTCAGCTGTGTTCTTTGGTGCCTATGGCGGTGGGCAAGTCAGTTTCATTTCTCTGAC	2991
QY	1000	TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu	1019
DB	2992	TATGCCAAAGCCAAATATACAGGCCACATCATCATCATTTGATAAAAAACCCCTTTC	3051
QY	1020	IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe	1039
DB	3052	ATTGACAGCTACAGCACGGAAGGCCTTAATGCCGAACACATTGGAAGGAATGTCACTTT	3111
QY	1040	AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer	1059
DB	3112	GGTGAAGTTGTATTCAACTATCCACCACCGACCATCCAGTGCTTCAGGAGCATGAGC	3171
QY	1060	LeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer	1079
DB	3172	CTGGAGGTGAAGAAGGCCAGACGCTGGGTCTGGTGGCAGCAGTGGCTGTGGGAAGAGC	3231
QY	1080	ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp	1099
DB	3232	ACAGTGGTCCAGCTCTCGAGCGGTTCTACGACCCCTTGGCAGGGAAGTGCTGCTGAT	3291
QY	1100	GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer	1119
DB	3292	GGCAAGAAATAAAGCGACTGAATGTTTCAGTGGCTCCGAGCACACCTGGGCATCTGTGCC	3351
QY	1120	GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer	1139
DB	3352	CAGAGGCCATCTGTTTGACTCGACGATTTGTGAGAACAATCCCTATGAGACACACAGC	3411
QY	1140	ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe	1159
DB	3412	CGGGTGTGTACAGGNAGAGATCGTGAGGGCAGCAAAAGGAGGCCAATACATGCTTTC	3471
QY	1160	IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspIysGlyThrGlnLeuSer	1179
DB	3472	ATCAGATCACTGCTTAATAAATATAGCACATAAGTAGGACAAAGAACTCAGCTCTCT	3531
QY	1180	GlyGlyClnIysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu	1199
DB	3532	GGTGGCCAGAAACAACCATTTGCCATAGCTCGTGGCCCTGTTAGACAGCCCTCATATTTTC	3591
QY	1200	LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla	1219


```
Db 3592 CTTTGGATGAAGCCACGCTAGCTCTGGATACAGAAAGTGAAGGTTGCCAAGAGCC 3651
|||||
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 3652 CTGGACAAAGCCAGAGAGCCGACCTGCATGTGTGATGCTCACCCTGTCCACCATC 3711
|||||
Qy 1240 GlnAsnAlaSerLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||
Db 3712 CAGAAATGCAGACTTAATAGTGGTGTTCAGAAAGTGGCAGAGTCAAGGAGCATGCACGCAT 3771
|||||
Qy 1260 GlnGlnLeuAlaGlnLysGlyIleThrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 3772 CAGCAGCTGCTGGCAGCAAGAAAGGCATCTATTTTCAATGGTCAAGTCCAGGCTGGAACA 3831
|||||
Qy 1280 LysArg 1281
|||||
Db 3832 AAGCGC 3837

RESULT 14
LOCUS AX336420
DEFINITION Sequence 6929 from Patent WO0194629.
ACCESSION AX336420
VERSION AX336420.1 GI:18127139
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 6929 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Source
1..4646
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1371 a 892 c 1129 g 1254 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 4646
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.47% Indels: 4
DB: 6 Gaps: 4

US-09-672-725c-23 (1-1281) x AX336420 (1-4646)

Qy 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
|||||
Db 425 ATGGATCTTGAAGGGACCGCAATGGAGGACCAAGACAGAGAACTTTTAAACTGAAC 484
|||||
Qy 20 LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
|||||
Db 485 AATAAAGTGAAGAAA---GATAAGAAAGAAAGAAACCAACTGTCAGTGTATTTCATG 541
|||||
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
|||||
Db 542 TTTCCCTATTCAAAATGGCTTGACAAAGTGTATATGGTGGGGAACCTTGGCTGCCATC 601
|||||
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspPhe 79
|||||
Db 602 ATCCATGGGGCTGGACTTCCCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 661
|||||
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
|||||
Db 662 GCAAAATGCAGGA---AATTTAGAAGATCTGATGTCAAACATCATCAATAAGAGTGATATC 718
|||||
```

```
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
|||||
Db 719 AATGATACAGGTTCTTCATGAAT---CTGGAGGAAGACATGACCATGCTTATAT 775
|||||
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
|||||
Db 776 TACAGTGGAAATGGTCTGGGCTGCTGCTGCTTACATTCAGGTTTCATTTGGTGC 835
|||||
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
|||||
Db 836 CTGGCAGCTGGAAGACAAATACACAAAATAGAAAACAGTTTTTTCATGCTATAATGCGA 895
|||||
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
|||||
Db 896 CAGGAGATAGCTGGTGTGGTGGAGCTTAAACCCGACTTACAGAT 955
|||||
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
|||||
Db 956 GATGCTCTTAAGATTAAATGAAGTTATTGGTGACAAAATTTGGAATGTTCTTCAGTCAATG 1015
|||||
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
|||||
Db 1016 GCACATATTTTCACTGGGTTTATAGTAGATTTACACGTGTTGGAAGCTAACCCCTTGTG 1075
|||||
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
|||||
Db 1076 ATTTTGGCCATCAGTCTGTTCTGGAGCTGCAGCTGCTGGGCAAGATGACTATCT 1135
|||||
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
|||||
Db 1136 TCATTTACTGATAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 1195
|||||
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
|||||
Db 1196 TTGCAGCAATAGCACTGATTGCAATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 1255
|||||
Qy 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
|||||
Db 1256 AACAAAAATTTAGAAAGAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCAATATT 1315
|||||
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
|||||
Db 1316 TCTATAGGTGCTGCTTCTCTGATGATGATGCTTATGCTCTGGCTTCTGGTATGGG 1375
|||||
Qy 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
|||||
Db 1376 ACCACTTGGTCTCTCAGGGGAATATTTATTGGCAAGTACTCTACTGTATTCTTTCT 1435
|||||
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
|||||
Db 1436 GTATTAAATTTGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAT 1495
|||||
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
|||||
Db 1496 GCAAGAGGAGCAGCTTATGAATCTTCAAGATAATTGATAATAAGCAAGTATTGACAGC 1555
|||||
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
|||||
Db 1556 TATTGGAAGAGTGGCACAACACAGATAATATTAAAGGAATTTGGAATTCAGAAATGTT 1615
|||||
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
|||||
Db 1616 CACTTCAAGTTACCCATCTCGAAAAGAAAGTTAAGATCTTGAAGGCGCTGAACCTGAAGGTG 1675
|||||
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAspSerGlyCysGlyLysSerThrVal 439
|||||
Db 1676 CAGAGTGGGCAGACGGTGGCCCTGGTGTGAAACAGTGGCTGTGGGAAGAGACACACAGTC 1735
|||||
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
|||||
Db 1736 CAGCTGATGCAGAGGCTCTATGACCCCAACAGAGGGGATGGTCAGTGTTCATGGACAGGAT 1795
|||||
```


QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
|||||
Db 1796 ATTAGGACCATAAATGTAAGGTTCTACGGGAAATCATTTGGTGTGGTGTAGTCAGGAACCT 1855
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
|||||
Db 1856 GTATTGTTGCCACCAGTAGCTGAACAATTCGCTATGGCCGTGAATAATGTCAACATG 1915
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
|||||
Db 1916 GATGAGATTGAGAAAGCTGCAAGAGCAATGCCATGACATTTATCATGAACCTGCCT 1975
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyValAlaGlnLeuSerGlyGlyGlnLysGln 539
:::|||||
Db 1976 CATAAATTTGACACCCCTGGTTGGAGAGAGAGGGCCAGTTGAGTGGTGGGCGACAGACAG 2035
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
|||||
Db 2036 AGGATCGCCATTGCACGTGCCCTGGTTGCGCAACCCCAAGATCCTCCTGCTGGATGAGGCC 2095
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
|||||
Db 2096 ACGTGACCCCTGGACACAGAAAGCGAAGCAGTGGTTCCAGTGGCTCTGGATAAGGCCAGA 2155
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
|||||
Db 2156 ANAGCTGGACCACCATTTGTATGATAGCTCATCGTTTGTCTACAGTTCTGTAATGCTGACGTC 2215
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 2216 ATCGCTGGTTTCGATGATGAGTCAATTTGTGAGAAAGAAATCATGATGAACATCAAGAA 2275
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
|||||
Db 2276 GAGAAAGCAATTAATCTCAAACTTGTCCAAATGCAGACAGCAGGAAATGAAGTTGAATTA 2335
QY 640 GluAsnAlaThrGlyLysSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
|||||
Db 2336 GAAATGCAGCTGATGAATCCAAAGTGAAATTTGATGCCCTTGGAAATGTCTTCAAAATGAT 2395
QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
|||
Db 2396 TCAAGATCCAGCTAAATAAGAAAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAAGCC 2455
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
|||||
Db 2456 CAAGACAGAAAGCTAGTACCAAAAGAGGCTCTGGATGAAAGTATACCTCCAGTTTCCCTT 2515
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
|||||
Db 2516 TGGAGGATTATGAGCTAAATTTAACTGAATGGCTTATTTTGTGTTGGTGTATTGTTGT 2575
QY 720 AlaIleIleAsnGlyLysGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
|||||
Db 2576 GCCATTATAAATGGAGGCTCAACACAGCATTTCGAATAATATTTTCAAAAGATTATAGGG 2635
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
:::|||||
Db 2636 GTTTTACAGAAATTGATGATCCCGAAACAAACAGACAGATAGTAACTGTTGTTTCACTA 2695
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
|||||
Db 2696 TTGTTTTCAGCCCTGGGAATTTATTTCTTTTATACATTTTTCCTTCAGGGTTTTCACATTT 2755
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
|||||
Db 2756 GGCAAAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATGTTGTTTCCGATCCATGCTC 2815
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
|||||
Db 2816 AGACAGGATGTGAGTTGTTGATGACCTTAAACACCATCTGGAGCATTTGACATACCAGG 2875
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleIleGlySerArgLeuAlaValIleThr 839
|||||

Db 2876 CTCGCCAATGATGCTGCATCAAGCTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTAATTACC 2935
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleTyrGlyTrpGlnLeu 859
|||||
Db 2936 CAGAATATAGCAAAATCTTGGGACAGAAATATATATATCTTCATCTATGTTGGTGGCAACTA 2995
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
|||||
Db 2996 ACACGTGTACTCTTAGCAATTTACCCATCATTTGCAATAGCAGAGATTTGTTGAATGAA 3055
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 899
|||||
Db 3056 ATGTTGCTGGAACAAGCACTGAAGATAAGAAAGAACTAGAAAGTCTCGGGAAGATCGCT 3115
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
:::|||||
Db 3116 ACTGAAGCAATAGAAAACCTTCGAAACCGTGTCTTTGACTCAGGAGCAGANGTTGAA 3175
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
:::|||||
Db 3176 CATATGTATGCTCAGAGTTTGCAGGTACCATACAGAACTCTTTGAGGAAAGCACACATC 3235
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 3236 TTTGGAATTAACATTTTCCCTTCACCCAGGCAATGATGATATTTTTCCTATGCTGGATGTTTC 3295
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
:::|||||
Db 3296 CGGTTTGGAGCCCTTGTGTGGCACATAAACTCATGAGCTTTGAGAGATGTTCTGTAGTA 3355
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
:::|||||
Db 3356 TTTTCAGCTGTTGCTTTTGTGCCATGGCCGTGGGCAAGTCAGTTCAITTTGCTCTGAC 3415
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
:::|||||
Db 3416 TATGCCAAAGCCAAAATATACAGAGCCCATCATCATCATCATTTGAAATAACCCCTTGG 3475
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
:::|||||
Db 3476 ATTGACAGCTACAGCAGGAGCCCTTAATGCCGACACACATTTGGAAGAAATGTCACATTT 3535
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
:::|||||
Db 3536 GGTGAAGTTGATTTCAACTATCCACCCGACCGGACATCCAGTGTCTCAGGAGACTGAGC 3595
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
:::|||||
Db 3596 CTGGAGGTGAAGAAAGGCCAGACGCTGGCTCTGGTGGGCGAGCAGTGGCTGTGGGAAGAGC 3655
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
:::|||||
Db 3656 ACAGTGTCTCAGCTCTCTGGAGCGGTCTACGACCCCTTGGCAGGAAAGTGTGCTGTAT 3715
QY 1100 GlyLysGluLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
:::|||||
Db 3716 GGCAAGAAATAAAGGACTGGAATTTGCTAGTGGCTCCGACACACACCTGGGCATCTGTGTC 3775
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
:::|||||
Db 3776 CAGGAGCCCATCTGTTGACTCGACGATTTGCTGAGAACTATGCCCTATGGAGACACACAG 3835
QY 1140 ArgValValSerHisGluIleMetGlnAlaAlaLysGluAlaAsnIleHisIlePhe 1159
:::|||||
Db 3836 CGGGTGTGTACAGGAGAGATCTGTGAGGCGACGAAAGAGGCCAACATACATGCTCTTC 3895
QY 1160 IleLeuThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
:::|||||
Db 3896 ATCGAGTCACTGCTTAATAAATATAGCACTAAAGTAGAGACAAAGAACTCAGCTCTCT 3955
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
|||||

```
Db 3956 GGTGGCCAGAAACACCCATTGCCATAGCTCGTGGCCCTGTTAGACAGCCTCATATTTTG 4015
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGGATGAAGCCAGCTGAGCTGGATACAGAAAGTGAAGAGTTGTCACCAAGGCC 4075
Qy 1220 LeuAspLysAlaArgGluGluArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGGACAAACCCAGAGAAAGCCACCTGCATTGTGATTGCTCACCGCCTGTCCACCATC 4135
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGATGCACACTTAATAGGGTGTTCAGATGGCAGATGCAAGAGATGGCAGCAT 4195
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGTCGCACAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 4255
Qy 1280 LysArg 1281
Db 4256 AAGCGC 4261

RESULT 15
AX336708
LOCUS AX336708 4646 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 7217 from Patent WO0194629.
ACCESSION AX336708
VERSION AX336708.1 GI:18127427
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Young,P.E., Augustus M., Carter,K.C., Ebner,R., Endress,G.,
Horrihan,S., Soppet,D.R. and Weaver,Z.
Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 7217 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source 1. 4646
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1371 a 892 c 1129 g 1254 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 4646
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.47% Indels: 4
DB: 6 Gaps: 4

US-09-672-725C-23 (1-1281) x AX336708 (1-4646)
Qy 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 425 ATGGATCTTGAAGGGACCGCAATGGAGGAGCAAGAAAGAACTTTTAAACTGAC 484
Qy 20 LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 485 AATAAAAGTGAAAA---GATAAGAGGAAAAAACCACCACTGTCAGTGTATTTCATG 541
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 542 TTTCGCTATTCAAATGGCTTCACAAAGTTGATATGGTGGGAACCTTGGCTGCCATC 601
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 602 ATCCATGGGCTGGACTTCCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 661
```

```
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 662 GCAAAATGCGAGGA---AATTAGAAAGATCTGATGTCACACATCACTAATAGAAGTCATATC 718
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrThrThrThrThrThr 119
Db 719 AATGATACAGAGGTCTCTCATGAAT---CTGGAGGAAGACATGACAGGTATGCTCATATAT 775
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 776 TACATGGAATGGTGGCTGGGTGCTGCTGCTACATTCAGGTTTCATTTTGGTGC 835
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 836 CTGGCAGCTGGAAGACAAATAACAAAATTAGAAAACAGTTTTTTCATGCTATTAATGCGA 895
Qy 160 GlnGluIleGlyTrpPheAspValHisaspValGlyGluLeuLeuAsnThrArgLeuThrAsp 179
Db 896 CAGGAGATAGGCTGGTTTGTGATGTCACGATGTTGGGAGCTTAACACCCGACTTTACAGAT 955
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 956 GATGCTCTAGATTAATCAAGATTATGGTGACAAAATGGAAATGTTCTTCAGTCAATG 1015
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 1016 GCAACATTTTTCACCTGGGTTTATAGTAGGATTTACACGTGGTGTGAAAGCTTAACCTTTGTG 1075
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 1076 ATTTGGGCCATCAGCTCCTGTTCTTGGACTGTGACGTGCTGTCTGGGCAAGATACTATCT 1135
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 1136 TCATTACTGATAAAGAACTCTTAGCGTATGCCAAAGCTGGAGCAGTGTGAGAGGCTC 1195
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 1196 TTGGCAGCAATTAGAACTGTGATTGCATTTGGAGGACAAAAGAAAGAACTTGAAGGCTAC 1255
Qy 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle 299
Db 1256 AACAAAAATTAGAAAGAAAGCTAAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCAATATT 1315
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrTrpGly 319
Db 1316 TCTATAGTGTGCTCTTCTGCTGATCTATGCACTTATGCTCTGGCCTTCTGGTATGGG 1375
Qy 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 1376 ACCACCTTGGTCTCTCAGGGGAATATTCTATTGGACAAGTACTCTACTGTATTCTTTCT 1435
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1436 GTATTAAATGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAAAT 1495
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1496 GCAAGAGAGCAGCTTATGAATCTTCAAGATTAATGATTAATGAACCAAGTATTGACAGC 1555
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1556 TATTGGAAGAGTGGGACAAACACAGATAATATTAAAGGAAATTTGGAATTCAGAAATGTT 1615
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1616 CACTTCAAGTACCACATCTCGAAAAGAAAGTAAAGATCTTGAAGGGCCTGACCTGAAGGTG 1675
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1676 CAGATGGGACAGCGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAGAGACACACAGCTC 1735
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
```


QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
 Db 3956 GTGGCCAGAACACACGATGGCATAGCTGGCCCTGTAGACAGCCTCATATTGG 4015
 QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluValValGlnGluAla 1219
 Db 4016 CTTTGGATGAAGCCACGCTCAGCTCGGATACAGAAAGTGAAGGTGTCCAAGAGCC 4075
 QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
 Db 4076 CTGGACAAAGCCAGAGAAGCGCACCTGCATTGTGTGCTCACCCTGTCCACCATC 4135
 QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
 Db 4136 CAGAAATGCAGACTTAATAGTGTGTTTCAGAAATGCAGAGTCAAGAGCATGGCAGCAT 4195
 QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyPheSerMetValSerValGlnAlaGlyAla 1279
 Db 4196 CAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCATGGTCAGTGTCCAGGCTGGACA 4255
 QY 1280 LysArg 1281
 Db 4256 AAGCGC 4261

Search completed: November 7, 2002, 04:59:36
 Job time : 4140.81 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:36:49 ; Search time 301.645 Seconds
(without alignments)
7291.243 Million cell updates/sec

Title: US-09-672-725C-23

Perfect score: 6474

Sequence: 1 MDPEGRKGSAEKNEWKMGK.....LLAQKGIYFMSVQVQAKR 1281

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=Cgn2_1/USPTO_spool/US09672725/runat_04112002_124338_17877/app.query.fasta_1.6435
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LIST=45 -UNITS=bits -QFMT=fastap -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=ptc -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-USER=US09672725.ecgn_1_1_0_erunat_04112002_124338_17877 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	6474	100.0	4279	22	Dog P-glycoprotein
2	6469	99.9	4279	22	Dog P-glycoprotein
3	6468	99.9	4279	22	Dog P-glycoprotein
4	6461	99.8	4279	22	Dog P-glycoprotein
5	6425.5	99.3	4317	22	Dog P-glycoprotein
6	5866	90.6	3860	21	Human wild-type mu
7	5866	90.6	3860	24	Human BCRP DNA rel
8	5862	90.5	4186	22	Cynomologous monke
9	5859.5	90.5	4195	22	Cynomologous monke
10	5857	90.5	3860	21	Human G185V mutant
11	5857	90.5	3860	24	Human BCRP DNA rel
12	5857	90.5	4349	22	Human intestine ce
13	5857	90.5	4646	21	Human ATP binding
14	5857	90.5	6505	17	Hybrid vector pSF-
15	5857	90.5	8630	21	Retroviral M4 mdr-
16	5857	90.5	8630	21	Retroviral vector
17	5854	90.4	4669	8	Sequence of human
18	5851	90.4	4378	11	Multidrug Resistan
19	5849.5	90.4	3988	21	Human MDR-1 DNA.
20	5844	90.3	4646	15	Human multidrug re
21	5840	90.2	4669	14	Sequence of human
22	5830	90.1	4669	19	Human P glycoprote
23	5815.5	89.8	4264	19	Mutated human P-gl
24	5815.5	89.8	4264	19	Mutated human P-gl
25	5641	87.1	4788	21	Murine multidrug r
26	5641	87.1	4788	24	Mouse BCRP DNA rel
27	5621	86.8	4425	21	Rat multidrug resi
28	5611	86.7	4369	21	Rat multidrug resi
29	5301	81.9	4189	21	Murine multidrug r
30	5301	81.9	4189	24	Mouse BCRP DNA rel
31	5301	81.9	4313	14	Mouse multidrug re
32	5291.5	81.7	4233	21	Rat mdrlb2 (multis
33	5291.5	81.7	4233	22	Rat mdrlb2 multidr
34	4910.5	75.8	3924	21	Human ATP binding
35	4910.5	75.8	3924	21	Human MDR-3 DNA.
36	3628	56.0	2726	15	Multidrug-resistan
37	3628	56.0	2726	15	Multidrug-resistan
38	3628	56.0	2726	15	Multidrug-resistan
39	3305	51.1	4776	21	Human ATP binding
40	2833.5	43.6	4175	20	H. contortus PGP-A
41	2693	41.6	3942	23	Drosophila melanog
42	2555	39.5	4157	23	Drosophila melanog
43	2512	38.8	4047	20	cDNA encoding the
44	2491.5	38.5	4111	23	Drosophila melanog
45	2482	38.3	4002	20	cDNA encoding mult

ALIGNMENTS

RESULT 1

ID AAD03504

AD AAD03504 standard; cDNA; 4279 BP.

XX AAD03504;

XX 13-JUN-2001 (first entry)

XX Dog P-glycoprotein (PGP) allelic variant (Genotype A) cDNA.

XX Dog; P-glycoprotein allelic variant; multidrug transporter; MDR1;

XX drug bioavailability; transgenic animal; genetic model; ss.

OS Canis familiaris.

XX Key Location/Qualifiers

XX CDS 17..3862

XX /*tag= a

Db 1337 CTGATGCAGAGGCTCTATGACCCACAGATGGCATGGCTGTGATTGATGGACAGCATTT 1396
Qy 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db 1397 AGGACCAATAATGTAAGGCATCTTCGGGAATTAATCTGGTGGTGAGTCAGGAGCCCTGTG 1456
Qy 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1457 TTGTTTGCACCCAGATAGCTGAACACATTCGCTATGGCCGCGAAATGTCCACATGGAT 1516
Qy 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1517 GAGATTGAGAAGCTGTGAAGAGGCCAATGCCCTATGATTTTATCATGAACACTACCTAAT 1576
Qy 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
Db 1577 AAATTTGCACACTCTGGTTGGAGAGAGAGGGCCCGAGCTGAGTGGTGACAGAAACAGAGA 1636
Qy 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
Db 1637 ATCGCCATTGCTGGGCCCTTGGTTGCGCAACCCCAAGATTCTTCTGCTGGATGAGCAACG 1696
Qy 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1697 TCAGCTCTGACACTGAAATGGAAGCAGTGGTTTCAGGTGGCCCTGGATGAAGCCAGAAAA 1756
Qy 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1757 GGCGGAGTACCATTGTGATAGCTCATCGTTGTCTACAGTTGCTAATGCCGATGTCATT 1816
Qy 601 AlaglyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1817 GCTGGTTTGTATGATGAGTCAATGTGGAGAAAGAAATCATGATCAATCAATGAAAGAG 1876
Qy 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1877 AAGGCATTACTTCAAACTTGTCACATGCAGACAGAGAAATGAAATGAGTTAGTAA 1936
Qy 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1937 AATGCCACTGGTGAATCCAAAAGTGAAGTGAATGATGCCCTTGGAAATGTCTCCAAAAGATTCA 1996
Qy 661 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db 1997 GGGTCAGTTTATAAAAAGAGATCAACTCGCAGAGATATACATGCACCAAGGCCAA 2056
Qy 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
Db 2057 GACAGAAAAGCTGGTACAAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCTCTCTGG 2116
Qy 701 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db 2117 AGGATTCTGAAGCTGAACCTCAACTGAATGGCCTTATTTTGGTTGGTATATTTTGTGCT 2176
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2177 ATTATAACGGAGGGCTGCAACACAGCATTTTCAATAATATTTTCAAGGATTATAGGGATC 2236
Qy 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2237 TTTACCCGAGATGAGATCTCTGAACAAACACAGAGAATAGTATGATGTTTCTGTATTG 2296
Qy 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2297 TTTCTAGTCTCTGGAATATTTCTTTTATACATTTTCTCCAGGGCTTCACATTTGCG 2356
Qy 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAAGCTGGGGAGATCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTGAGA 2416
Qy 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGGATGTCAGCTGGTTGTATGACCCCTAANAACACCACTGGAGCATTTGACAAACCGGCTT 2476

Qy 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2477 GCCAATGATGCGGCTCAAGTTAAAGGGGTATAGTGTCCAGGCTTGCTGTCAATTACCCAG 2536
Qy 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
Db 2537 AATATAGCAAAATCTGGGACAGCATATATATCTTAATCTATGTTTGGCAATTAACA 2596
Qy 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTACTCTTAGCAATTTGACCATTTGCAATAGCAGGAGTTGTTGAAATGAAATG 2656
Qy 881 LeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
Db 2657 TTGCTGGGACAAGCACTGAAGATAAGAAAGAGCTAGAAGGAGCTGGGAAGATTCTCTACA 2716
Qy 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2717 GAAGCCATCGAAAACCTTCGGAACCTGTTGTTCTTTGACTCGGACGACAGAATTTGAATAC 2776
Qy 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2777 ATGTATGCACAGAGTTTGCAGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTTC 2836
Qy 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2837 GGGGTCTCATTTTCTATCACCAGCAATGATGATTTTCTATGCTGGCTGTCTCCGG 2896
Qy 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
Db 2897 TTTGTTGCTACTCTTGGTGGCAATCAGTTTCATGAACCTTCAGGATGTTCTTTGGTATTC 2956
Qy 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
Db 2957 TCAGTATATGTTTGGTGCCATGGCAGTGGGCGAGTCACTTCAATTTGCTCTGACTAT 3016
Qy 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
Db 3017 GCCAAGCCAAAGTATCAGCAGCCACGTCATCATGATCATTGAAGAAAGCCCTGATT 3076
Qy 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3077 GACAGCTACAGCCCTCAGCGCTCAAGCCAAATACGTTGGAAGAAATGTGACATTTAAT 3136
Qy 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3137 GAGGTGCTGTTCAACTATCCCACTCGACACATCCCGCTGCCAGGGCTCAGCCTC 3196
Qy 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
Db 3197 GAGGTGAAGAGGGCCAGCCCTGGCCCTCTAGTAGTACGCTGGCTGGGGAAGAGCAC 3256
Qy 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
Db 3257 GTTGTTCAGCTCCTAGAGCGCTTCTATGACCCCTTGGCTGGTTCAGTGTCAATTTGATGC 3316
Qy 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120
Db 3317 AAAAGATAAAGCACTGAATGTCTCAGTGGCTCCGACACACCTGGGCACTGCTGTCTCAG 3376
Qy 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
Db 3377 GAGCCCATCTGTTGACTGCAGCATTTGCCAGAACATTTGCCATTGAGAGAACAGCCGG 3436
Qy 1141 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1160
Db 3437 GTCGTATCATGAAGAGATTATGACGCGCAGCAAGGAGGCGCAACATACACCACTTCATC 3496
Qy 1161 GluThrLeuProGluLysTyrAsnThrArgValIcLysAspLysGlyThrGlnLeuSerGly 1180
Db 3497 GAGACACTCCTCGAATAATACACACAGAGTAGGAGACAAAGAAACCCAGCTCTCTGTT 3556

QY 1181 GlyInLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
DB 3557 GGCAGAAACAGCGATTGCCATAGCTCGCGCTCTGTGTAGACAGCCCTCATATTTTGCTT 3616
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
DB 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAGAAGCCCTG 3676
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
DB 3677 GACAAAGCCAGAGAGCGCACCTGCTATGTGATCGGCCACCGCTGTGCCACCATCCAG 3736
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
DB 3737 AATGCAGATTATAATAGTGTGTTCAGAAATGGCAAGTCAAGAGCATGGCCACATCAA 3796
QY 1261 GlnLeuLeuAlaGlnLysGlyIleThrPheSerMetValSerValGlnAlaGlyAlaLys 1280
DB 3797 CAGCTGCTGGCCCAAGAGGCATCTATTTTCCATGGTCAGTGTCCAGGCTGGAGCAAG 3856
QY 1281 Arg 1281
DB 3857 CGC 3859
RESULT 2
AAD03488
ID AAD03488 standard; cDNA; 4279 BP.
AC AAD03488;
XX 13-JUN-2001 (first entry)
DT Dog P-glycoprotein (PGP) cDNA #1.
DE Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW drug bioavailability; transgenic animal; genetic model; ss.
XX Canis familiaris.
FH Key Location/Qualifiers
FT CDS 17..3862
FT FT /*tag= a
FT FT /product= "Dog P-glycoprotein (PGP) #1"
XX WO200123540-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26767.
XX 28-SEP-1999; 99US-0156510.
XX (GENT-) GENTEST CORP.
XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX WPI: 2001-235373/24.
XX P-PSDB; AAE00303.
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX for determining the bioavailability of drugs and for screening for dog
XX PGP inhibitors -
XX Claim 3; Page 58-63; 11lpp; English.
XX The invention relates to dog P-glycoprotein (PGP) also referred
XX as multidrug transporter (MDR1) and nucleic acids encoding them.
XX The invention also includes fragments and biologically functional
XX variants of dog P-glycoprotein. PGP and their nucleic acids are
XX useful for determining the bioavailability of drugs and for
XX screening PGP inhibitors. They are useful for the diagnosis and
XX treatment of conditions characterised by PGP activity, by
XX reducing or increasing PGP activity in a cell. PGP nucleic acids

CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) cDNA. This
CC sequence is also referred as Genotype C cDNA. The
CC PGP enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.

XX Sequence 4279 BP; 1294 A; 834 C; 1008 G; 1143 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6469.00 Matches: 1280
Percent Similarity: 99.92% Conservative: 0
Best Local Similarity: 99.92% Mismatches: 1
Query Match: 99.92% Indels: 0
DB: 22 Gaps: 0

US-09-672-725C-23 (1-1281) x AAD03488 (1-4279)

QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
DB 17 ATGGATCCTGAAGAGGCCGTAAGGGGAGTGCAGAGAAAGAACTTCTGAAAAATGGGCAAA 76
QY 21 LysSerLysLysAsnGluLysGluLysLysProThrValSerThrPheAlaMetPhe 40
DB 77 AAAGTAAAAAATGAGAAGAAAGAAACCACTGTCCAGCAGCTTTGGCAATGTTT 136
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 60
DB 137 CGCTATTCAAATGGCTTGATAGTTGTATATTTGGTGGGCAATGGCTGCCATCATC 196
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
DB 197 CATGAGCTGCACCTCCCTCATGATGCTGGTTTGGAAACATGACAGATAGCTTTGCA 256
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
DB 257 AATGCAGGAATTTCAAGAAACAAAACCTTTCCAGTTATAATTAATGAAGTATTACGAAC 316
QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120
DB 317 AATACACAACTTTTCATCAACCATCTGGAGGAGAAATGACACATGCTATCCCTATTATTC 376
QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaIleValIleGlnValSerPheTrpCysLeu 140
DB 377 AGTGGATCGGTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 436
QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
DB 437 GCAGCAGGAAGACAGATACATAAAATAGAAACAAATTTTTCATGCTATCATGGCAGCAG 496
QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
DB 497 GAGATTGGCTGTTTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 556
QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
DB 557 GTCTCCAAAATCAATGAAGGAATTGGCGACAAAATTTGGAATGTTCTTTCACATCAATAGCA 616
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
DB 617 ACATTTTTCACCGGTTTATAGTGGGTTTACCGTGGTGGAGCTTAACCCCGCTCACAGACCAT 676
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaIleIleTrpAlaLysIleLeuSerSer 240
DB 677 TTGGCCATCAGCCCTGTTCTTGGACTTTTCAGCCGCATCTGGGCAAGAAAGTACTATCTTCA 736
QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260

Db 737 TTTACTGATAAGAACTCTTGGCCCTATGTCACAAAAGCTGGAGCAGTGTGAGAAAGTCTTA 796
Qy 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
Db 797 GCACCAATCAGAAGCTGTGATGGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAAC 856
Qy 281 LysAsnLeuGluGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIleSer 300
Db 857 AAAAATTAGAAAGCTAAGGAATTGGGATAAAGAAAGCTATCACGGCCAACTTCT 916
Qy 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrPyrGlyThr 320
Db 917 ATTGGTCGGCTTTCTATTGATCATGCTATCATATGCTCTGGCTTTCTGTATGGACC 976
Qy 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
Db 977 TCCITGGTCCCTCCAGTAATATCTATTGGCAAGTACTCACTGCTCTTCTTTCTGTGA 1036
Qy 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db 1037 TTAATTGGGGCTTTTAGTATTGGACAGCATCCCCAGCATTTGAAGCATTTGCAACGCA 1096
Qy 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
Db 1097 AGAGGAGCAGCTTATGAATCTTCAAGATAATTGACATAATAACCAAGCATTTGACAGTAT 1156
Qy 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1157 TCGAAGAGTGACATAAACCAAGATAATTAAGGGAATTTGGAATTCAAAATGTTTAC 1216
Qy 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1217 TTCAGTTACCTTCTCGAAGAAGAGTTAAGATCTTAAGGGTCTCAACCTGAAGTTAC 1276
Qy 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1277 AGTGGCAGACAGTGGCGCTGTTGGGAACAGTGGCTGCGGGAAGACACACCGTCAG 1336
Qy 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db 1337 CTGATGACAGAGGCTCTATGACCCACAGATGGCATGCTGTGTATGTAGACAGGACATT 1396
Qy 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db 1397 AGGACCAATAATGTAAAGCATCTTCGGGAATTTACTGGTGTGGTACGTACGAGCCCTGTG 1456
Qy 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1457 TTGTTTGCACACGATAGCTGAAACATTCGCTATGGCCGCGAAAATGTCAACATGGAT 1516
Qy 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1517 GAGATTGAGAAAGCTGTTAAGGAAGCAATGCCCTATGATTTATCATGAACACTACCTAAT 1576
Qy 521 LysPheAspThrLeuValGlyArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
Db 1577 AAATTTGNACACTCTGGTGGAGAGAGGGGCCAGCTGAGTGGTGGACAGAAACACAGA 1636
Qy 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
Db 1637 ATCGCCATTGCTCGGGCCCTGGTTCGCAACCCCAAGATTCTTCTGCTGGATGAGGCAACG 1696
Qy 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1697 TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCTGGATAAGGCCAGAAAA 1756
Qy 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1757 GCCCGGACTACCATTTGATAGTACATCGTTTGTCTACAGTTCGTAATGCCGATGTCTATT 1816
Qy 601 AlaGlyPheAspAspGlyValIleValIclulysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1817 GCTGGTTTGTATGATGGATCATTTGTGGAGAAAGAAATCATGATGAACATCATGAAGAG 1876

Qy 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1877 AAGGGCATTTACTTCAAACTTGTCACAATGCACACAGAGAAATGAAATTCAGTTAGAA 1936
Qy 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1937 AATGCCACTGGTGAATCCAAAAGTGAAGTGCCTTGGAAATGTCTCCAAAAGATTCA 1996
Qy 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db 1997 GGTCCAGTTTATAAAAAAGAGATCACTCGCAGAGATATACATGCACCAAGGCCAA 2056
Qy 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTyr 700
Db 2057 GACAGAAAGCTTGGTACAAAAGAGGACTTGAATGAGAATGTACCTCCAGTTCCTCTGG 2116
Qy 701 ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 720
Db 2117 AGGATTTCTGAAGCTGAACCTCAACTGAATGCCCTTATTTGTGTTGGTATATTTTGTGCT 2176
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2177 ATTATAACGGAGGCTGCAACCCAGCATTTTCAATAATATTTTCAAGGATTTATAGGGATC 2236
Qy 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2237 TTTACCCGAGATGAGATCCTGNAACAAAACGACAGAATAGTAACATGTTTCTGTATTG 2296
Qy 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2297 TTTCTAGTCTCTGGAATTTATTTCTTTTATTACATTTTCTCCAGGGCTTCACATTTGGC 2356
Qy 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAAGCTGGGAGATCCTCACTAAGCGGCTCGATACATGGTTTTCAGATCCATGCCAG 2416
Qy 801 GlnAspValSerThrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGATGTACGTGGTGTGTGATGACCTTAAACACCCACCTGGAGCATTGACACCCAGCTT 2476
Qy 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2477 GCCAATGATGCGGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTCTGCTCATTTACCAG 2536
Qy 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyThrPheGlnLeuThr 860
Db 2537 AATATAGCAAACTCTGGGACGACATTTATATATCTTAACTATGTTGGCAATTAACA 2596
Qy 861 LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTTACTCTTAGCAATGTGACCCATCATTGCATAGCAGGAGTGTGTAATGAAATG 2656
Qy 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
Db 2657 TTGCTGGACAACACTGAAAGATAAGAAAGAGCTAGAAGAGCTGGGAAGATTGTCTACA 2716
Qy 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2717 GAAGCCATCGAAAACCTTCGGAACCTGTTGTTCTTTGACTCGGAGCAGAGTTCGAATAC 2776
Qy 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2777 ATGTATGCACAGAGTTGCAAGTACCATTACAGAAACTCTTTTGAGGAAACACACATCTTC 2836
Qy 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2837 GGGGTCTCATTTCTATACCCAGGCAATGATGATATTTTCTCTATGCTGGCTGTTCCGG 2896
Qy 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 980
Db 2897 TTTGGTGCCTACTTGTGGCAATAGTTCATGAACATTTTCAGGATGTTCTTTTGTGTTATTC 2956

```
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
DB 2957 TCAGCTATTGCTTTGGGCCATGCGATGGGAGGTCAGTTTCATTTGCTCCGACATAT 3016
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
DB 3017 GCCAAGCCAAAGTATCAGCAGCCACGTCATCATGATCATTTGAAAAAGCCCTCGATT 3076
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
DB 3077 GACAGCTACAGCCCTCAGCGGCTCAAGCCAAATAGCTTGAAGGAAATGTGACATTTAAT 3136
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
DB 3137 GAGGTGCGTTCACATATCCCACTCGACGACACATCCCGTGGCTCCAGGGGCTGAGCCTC 3196
QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyLysSerThr 1080
DB 3197 GAGGTGAAGAGGGCCAGACGCTGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGAGCACA 3256
QY 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
DB 3257 GTTGTTCAGCTCCTAGAGCGCTTCATGACCCCTTGCGTGTGCTAGTCTAATTGATGGC 3316
QY 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120
DB 3317 AAAGAGATAAAGACCTGAATGTCAGTGGCTCGGAGCACACCTGGGGCATCTGCTCTCAG 3376
QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAsnSerArg 1140
DB 3377 GAGCCCATCTGTTGACTGCGCAGCATTCGCCGAGACATTTGCTATGGAGACAACAGCCGG 3436
QY 1141 ValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPheIle 1160
DB 3437 TCTGTATCACATGAAGATTATGCAGCAGCCCAAGGAGGCCAACATACACCACTTCATC 3496
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
DB 3497 GAGACATCTCCCTGAGAAATACAAACACAGAGTAGGAGACAAGGAACCCAGCTCTCTGGT 3556
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
DB 3557 GCCAGAAACAGCGCATTTGCCATAGCTCGCGCTCTTGTATGACAGCCCTCATATTTGCTT 3616
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
DB 3617 TTGGATGAAGCTACATCAGCTGCGATACAGAAAGTGAAGAGTTGTCCAAAGAAGCCCTG 3676
QY 1221 AspLysAlaArgGluGluArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
DB 3677 GACAAGCCAGAGAAGGCCGACCTGCGATTTGTATCGCCCGCTTGTCCACCATCCAG 3736
QY 1241 AsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
DB 3737 ATGTCAGATTTAATAGTGGTCTTCAGAAATGGCAAGTCAAGGAGCATGGCACACATCAA 3796
QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
DB 3797 CAGCTGCTGGCCAGAAAGGATCATTTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3856
QY 1281 Arg 1281
DB 3857 CCG 3859
RESULT 3
AAD03505
ID AAD03505 standard; cDNA; 4279 BP.
XX
AC AAD03505;
XX
DT 13-JUN-2001 (first entry)
XX
DE Dog P-glycoprotein (PGP) allelic variant (Genotype B) cDNA.
```

```
XX
KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
MDR1; drug bioavailability; transgenic animal; genetic model; ss.
XX
OS Canis familiaris.
XX
Key Location/Qualifiers
CDS 17..3862
FT /*tag= a
FT /product= "Dog P-glycoprotein (PGP) allelic variant
FT (Genotype B) protein"
FT replace (91, T)
FT /*tag= b
FT replace (607, C)
FT /*tag= c
XX
WO200123540-A2.
XX
PN
XX
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26767.
XX
XX 28-SEP-1999; 99US-0156510.
XX
XX (GENT-) GENTEST CORP.
XX
PA Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX
PI WPI; 2001-235373/24.
XX
DR P-PSDB; AAE00309.
XX
XX
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
XX Claim 9; Page 93-99; 11pp; English.
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
XX as multidrug transporter (MDR1) and nucleic acids encoding them.
XX The invention also includes fragments and biologically functional
XX variants of dog P-glycoprotein. PGP and their nucleic acids are
XX useful for determining the bioavailability of drugs and for
XX screening PGP inhibitors. They are useful for the diagnosis and
XX treatment of conditions characterised by PGP activity, by
XX reducing or increasing PGP activity in a cell. PGP nucleic acids
XX are used as oligonucleotide probes. Complements of PGP nucleic
XX acids are useful as antisense oligonucleotides, to induce a PGP
XX 'knockout' phenotype. They are used to prepare a non-human
XX transgenic animal, which are valuable as genetic models for
XX human diseases.
XX
XX The present sequence is dog P-glycoprotein (PGP) allelic variant
XX (Genotype B) cDNA. The PGP enzyme functions as an efflux pump
XX exporting small molecules across the cell membrane. This enzyme
XX is a member of the ABC transporter family.
XX
XX Sequence 4279 BP; 1296 A; 833 C; 1008 G; 1142 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6468.00 Matches: 1280
Percent Similarity: 99.92% Conservative: 0
Best Local Similarity: 99.92% Mismatches: 1
Query Match: 99.91% Indels: 0
DB: 22 Gaps: 0
US-09-672-725c-23 (1-1281) x AAD03505 (1-4279)
QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
DB 17 ATGATCTCTGAAGGAGCCGCTAAGGGAGTGCAGAGAAGAACTCTCGAAAAATGGCAAA 76
QY 21 LysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
|||||
```

Db 77 AAAAGTAAAAAAGAGAGAGAAAAAGAAACCAACCTGTCTAGCAGCGTTTGTCAATGTTT 136
 Qy 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 60
 Db 137 CGCTATTCAAAATGGCTTGATAGGTGTATATGTTGGTGGGACAATGGCTGGCATATC 196
 Qy 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
 Db 197 CATGGAGCTGCACCTCCCTCTCATGCTGCTGTTTTGGAAACATGACAGATAGCTTTGCA 256
 Qy 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
 Db 257 AATCAGGAATTTCAAGAAACAAAACCTTTCCAGTTATAATTAATGAAGTATTACGAAC 316
 Qy 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120
 Db 317 AATACACAACATTTTCATCAACCATCTGGAGGAGAAATGACCAGTATGCTTATATTAC 376
 Qy 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
 Db 377 AGTGGATCGGTGCTGGCTGCTGGTGGCTTTACATCCAGGTTTCATTTCTGGTGGCTG 436
 Qy 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
 Db 437 GCAGCAGGAACACAGATACTCAAAATTAGAAAACAATTTTTTCATGCTATCATGCGACAG 496
 Qy 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
 Db 497 GAGATTTGGCTGGTTGACGTGCATGACGTTGGGAGCTTAACACCCGGCTCACAGACGAT 556
 Qy 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
 Db 557 GTCTCCAAAATCATGAAGGAATGGCGACAAAATTTGGAATGTTCTTTCATCATAGCA 616
 Qy 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
 Db 617 ACATTTTTACCGGTTTTATAGTGGGTTTTACAGCTGGTTGGAAGCTTAACCCCTTGTGATT 676
 Qy 221 LeuAlaIleSerProValIleuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
 Db 677 TTGGCCATFACGCCCTGTCTTGGACTTTCAGCCGCCATCTGGGCAAGATACTATCTCA 736
 Qy 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
 Db 737 TTTACTGATAAAGAACTCTGGCCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGCTTA 796
 Qy 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
 Db 797 GCAGCAATCAGAACTGTGATTGCCCTTTGGAGGACAAAAGAAAGAACTTGAAGGTACAAC 856
 Qy 281 LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
 Db 857 AAAAATTTAGAAGAAGCTAAAGGAATTTGGGATAAAGAAAGCTATFACGGCCAACTTTCT 916
 Qy 301 IledGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
 Db 917 ATTGGTGGCGTTTCTTATTGATCTATGCATCATATGCTCTGGCTTCTGGTAGGGACC 976
 Qy 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
 Db 977 TCCTTGGTCTCTCCAGTGAATATTCATTGGACAAGTACTACTGCTCTTCTTTCGTGA 1036
 Qy 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
 Db 1037 TTAATTTGGGCGTTTATGATTTGGACAGCATCCCAAGCATTTGAAGCATTTGCAAGGCA 1096
 Qy 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
 Db 1097 ACAGGACAGCTTATGAATTCACAGATATTGACAAATTAACCAACCATTTGACAGCTAT 1156
 Qy 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
 Db 1157 TCGAAGATGGACATAAACACAGATAATATTAAAGGAATTTGGAATTCAAAATGTTTAC 1216

Qy 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
 Db 1217 TTCAGTATACCTTCTCGAAAAGAGTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTCA 1276
 Qy 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
 Db 1277 AGTGGGCACACAGTGGCGCTGGTTGGAAACAGTGGCTGGGGAAGACGACCGCTGCAG 1336
 Qy 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
 Db 1337 CTGATGCAGAGGCTCTATGACCCACAGATGGCATGGTCTGTATTGATGCAGACGACAT 1396
 Qy 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
 Db 1397 AGACCATTAATTAAGGCATCTTCGGGAAATTCAGTGTGGTGGTGCAGAGGCTGAG 1456
 Qy 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
 Db 1457 TTGTTTGCACACAGATAGCTGAAAACAATTCGCTATGGCGCGAAAATGTCACCATGGAT 1516
 Qy 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
 Db 1517 GAGATTGAGAAAGCTGTTAAGGAAGCCAATGCTATGATTTTATCATGAACACTACCTAAT 1576
 Qy 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
 Db 1577 AAATTTGACACTCTGTTGGAGAGAGGGGCCAGCTGAGTGGTGGAGCAGAAACAGAGA 1636
 Qy 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
 Db 1637 ATCGCCATTGCTCGGGCCCTGGTTCCCAACCCCAAGATTTCTTCTGCTGATGAGCAACG 1696
 Qy 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
 Db 1697 TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAGTGGCGCCCTGGATGAAGCCAGAAA 1756
 Qy 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
 Db 1757 GGGCGACATACCATNTGTATAGCTCATCGTTTGTCTACAGTTCGTAATGCCCATGTCATT 1816
 Qy 601 AlaGlyPheAspGlyValIleValGlyLysGlyAsnHisAspGluLeuMetLysGlu 620
 Db 1817 GCTGGTTTTGATGATGAGTCAATTTGGAGAAAGAAATCATGATGAACATCAAGAGAG 1876
 Qy 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
 Db 1877 AAGGGCATTTACTTCAAACTTGTCACAATGCAGACAAGAGAAATGAAATTCAGTTAGAA 1936
 Qy 641 AspAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
 Db 1937 AATGCCACCTGGTGAATCCAAAAGTGAAGTGAATGATGCCCTTGGAAATGCTCCAAAAGATTCA 1996
 Qy 661 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
 Db 1997 GGGTCCAGTTTAATAAAGAAAGATCACTCGCAGGAGTATACATGCACCAACAGGCCAA 2056
 Qy 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
 Db 2057 GACAGAAAAGCTTGGTCAAAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCCCTCTGG 2116
 Qy 701 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
 Db 2117 AGGATTTCTGAAGCTGAACCTCAACTGAATGGCCTTATTTTGTGGTGGTATATTTTGTGCT 2176
 Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
 Db 2177 ATTATAAACGGAGGCTTCACACAGCATTTTCAATAATATTTTCAAGGATTATAGGATC 2236
 Qy 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
 Db 2237 TTTATCCCGATGAGGATCCTGAAACAAAACGACAGAAATAGTAACATGTTTCTGTATTG 2296

```
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2297 TTTCTAGTCCCTGGGAATTAATTTCTTTATTACATTTTCTCCAGGGCTTCACATTTGGC 2356
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAAGCTGGGAGATCCTCACTAAGCGCTTCGATACATGGTTTTCAGATCCATCCTCGA 2416
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGGATGTCAGCTGGTTGTGTGAGCCCTAAAACACCACCTGGAGCATGTGACAAACAGGCTT 2476
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2477 GCCAATGATCGGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTCATTACCCAG 2536
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
Db 2537 AATATAGCAAAATCTTGGGACAGGCATTATATATCCTTAATCTATGGTTGGCAATTAA 2596
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTACTCTTAGCAATGTGACCATCATTTGCAATAGCAGGAGTTGTGAATGAAATG 2656
QY 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
Db 2657 TTGCTGGACAAGCACTGAAAGATAAAGAAAGAGTAGAAGGAGCTGGGAAGATTGCTACA 2716
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2717 GAAGGCATCGAAAACTTCGGAACCTGTTTCTTTGACTCGGGAGCGAAGTTTGAATAC 2776
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2777 ATGTATGCACAGAGTTTGCAGTACCACATACAGAACTCTTTGAGGAAGCACACATCTTC 2836
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2837 GGGGTCTCATTTTCTATCACCAGCAATGATGATATTTTCTATGCTGGCTGTTCCGG 2896
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
Db 2897 TTTGGTGCTACTTGGTGGCAATGAGTTTCATGAACTTTTCAGGATGTTCTTTTGGTATTC 2956
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
Db 2957 TCAGCTATGTCTTGGTGCCATGGCATGGCGAGGTCAGTTCTATTGCTCTCGACTAT 3016
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
Db 3017 GCCAAAGCCAAAGTATCAGCAGCCACGCTCATCATGATCATTTGAAAGAGCCCTCTGATT 3076
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3077 GACAGCTACAGCCCTACGCGCTCAAGCCAAATAGTTGGAAGGAATGTGACATTAAAT 3136
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3137 GAGTCTGTTCACATATCCCACTCGACACACATCCCGTGCTCCAGGGGCTGAGCCCTC 3196
QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
Db 3197 GAGGTGAAGAGGGGCGAGACGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGAGACACA 3256
QY 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
Db 3257 GTTGTTCAGCTCCTAGAGCGCTTCTATGACCCCTTGGCTGGCTGCTAAATTTGATGGC 3316
QY 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120
Db 3317 AAAGAGATAAAGCACCTGAATGTCAGTGGCTCCGAGCACACCTGGGCATGCTGCTCAG 3376
QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
```

```
Db 3377 GAGCCCATCCTGTTTGTACTGCAGCATTCGCGAGAAATTCCTTATGGAGACACACCGG 3436
QY 1141 ValValSerHisGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1160
Db 3437 GTCGTATCATAGAGAGATTATGCAGGAGCCAGAGAGCCCAACATACACACTTCATC 3496
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
Db 3497 GAGACACTCCTCGAATAATACACACAGAGTAGGAGACAAAGAACCCAGCTCTCTGCT 3556
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
Db 3557 GGCAGAAACAGCGCATATGCCATAGCTCGCGCTCTTTGTAGACAGCTCATATTTTGCTT 3616
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
Db 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAGNAGCCCTG 3676
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
Db 3677 GACAAAGCCAGAGAGCCGACCTGCATTGTATGATCGCCACCGCTTGTCCACCATCCAG 3736
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
Db 3737 AATGCAGATTTAATAGTGGTGTTCAGAAATGGCAAGTCAAGGAGCATGGCACACATCAA 3796
QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
Db 3797 CAGCTGCTGCCCCAGAAAGGATCTATTTTCCATGGTCAGTGTCCAGGCTGGAGCAAG 3856
QY 1281 Arg 1281
Db 3857 CGC 3859
RESULT 4
AAB03506
ID AAD03506 standard; cdNA; 4279 BP.
XX
AC AAD03506;
XX
DT 13-JUN-2001 (first entry)
XX
DE Dog P-glycoprotein (Pgp) allelic variant (Genotype D) cDNA.
XX
KW Dog; P-glycoprotein allelic variant; Pgp; multidrug transporter;
XX MDR1; drug bioavailability; transgenic animal; genetic model; ss.
XX
OS Canis familiaris.
XX
FH Key Location/Qualifiers
FT CDS 17..3862
FT FT /*tag= a
FT FT /product= "Dog P-glycoprotein (Pgp) allelic variant
FT FT (genotype D) protein"
FT FT replace (91, T)
FT FT /*tag= b
FT FT replace (607, C)
FT FT /*tag= c
FT FT replace (1001, T)
FT FT /*tag= c
FT FT replace (3458, A)
FT FT /*tag= c
XX
PN WO200123540-A2.
XX
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26767.
XX
PR 28-SEP-1999; 99US-0156510.
XX
PA (GENT-) GENTEST CORP.
```

Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 WPI: 2001-235373/24.
 P-PSDB; AAE00310.
 New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 for determining the bioavailability of drugs and for screening for dog
 PGP inhibitors -
 Claim 9; Page 102-107; 111pp; English.
 The invention relates to dog P-glycoprotein (PGP) also referred
 as multidrug transporter (MDR1) and nucleic acids encoding them.
 The invention also includes fragments and biologically functional
 variants of dog P-glycoprotein, PGP and their nucleic acids are
 useful for determining the bioavailability of drugs and for
 screening PGP inhibitors. They are useful for the diagnosis and
 treatment of conditions characterised by PGP activity, by
 reducing or increasing PGP activity in a cell. PGP nucleic acids
 are used as oligonucleotide probes. Complements of PGP nucleic
 acids are useful as antisense oligonucleotides, to induce a PGP
 'knockout' phenotype. They are used to prepare a non-human
 transgenic animal, which are valuable as genetic models for
 human diseases.
 The present sequence is dog P-glycoprotein (PGP) allelic variant
 (Genotype D) cDNA. The PGP enzyme functions as an efflux pump
 exporting small molecules across the cell membrane. This enzyme
 is a member of the ABC transporter family.
 XX SQ Sequence 4279 BP; 1296 A; 833 C; 1009 G; 1141 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 4279
 Score: 6461.00 Matches: 1278
 Percent Similarity: 99.92% Conservative: 2
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 99.80% Indels: 0
 DB: 22 Gaps: 0
 US-09-672-725c-23 (1-1281) x AAD03506 (1-4279)
 QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPhetrpLysMetGlyLys 20
 DB 17 ATGGATCCCTGAAGGAGCGCGTAAGGGAGTGCAGAGAAAGCACTTCTGGAAAATGGCGAAA 76
 QY 21 LysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
 DB 77 AAAAGTAAAAAAGAGAGAAAGAAAGAAAGAAACCAACTGTGACGACGTTTGCAATGTTT 136
 QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 60
 DB 137 CGCTATTCAAAATGGCTTGATAGGTGTATATGTTGGTGGGCAATGGCTGCCATCATC 196
 QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
 DB 197 CATGGAGCTGCACCTCCCTCTCATGATGCTGTTTGGAAACATGACAGATAGCTTTGCA 256
 QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
 DB 257 ATGCAGGAAATTCAGAAACAAAACCTTTCCAGTTATAATTAATGAAGTATTACGAAC 316
 QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120
 DB 317 AATACACAACATTTTCATCAACCATCTGGAGGAGGAAATGACCAGTATGCCCTATTATTAC 376
 QY 121 SerGlyIleGlyAlaGlyValLeuValAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 140
 DB 377 AGTGGGATCGGTGCTGGGGTGTGCTGCTTACATCCAGGTTTCATTCGTGGCTG 436
 QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
 DB 437 GCAGCAGGAGACAGATACTCAAAATTAGAAACAAATTTTTCATGCTATCATCGCAG 496

QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
 DB 497 GAGATTGGCTGGTTTGGACGTGCATGACGTTGGGAGCTTAACACCCGGCTCACAGACAT 556
 QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
 DB 557 GTCTCCAAATCAATGAAGGAATGGCGACAAATTTGGAATGTTCTTCAATCAATAGCA 616
 QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
 DB 617 ACATTTTTCACCGTTTATAGTGGGTTTACACGTTGGAGCTTAACCCCTTGTGATT 676
 QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
 DB 677 TTGGCCATCAGCCCTGTTCTTGGACTTTTCAGCCGCACTTGGGCAAGATACTACTTCA 736
 QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
 DB 737 TTTACTGATAAAGAACTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 796
 QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
 DB 797 GCAGCAATCAGAACTGTGATTGCCCTTGGAGCAGCAAAAGAAAGAACTTGAAGGTACAAC 856
 QY 281 LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
 DB 857 AAAAATTTAGAAAGAGCTAAAGGAATGGGATAAAGAAAGCTATCAGCGCCCAACATTCT 916
 QY 301 IleGlyAlaAlaPheLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrTyrGlyThr 320
 DB 917 ATTTGGTGGCTTCTTATTGATCTATGCATCATATGCTCTGGCTTCTGGTATGGGACC 976
 QY 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
 DB 977 TCCTTGGTCCCTCTCAGTGAATATACTATTGGACAAGTACTCACTGCTCTCTTTCTGTA 1036
 QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
 DB 1037 TTAATTTGGGGCTTTTAGTATTGGAGCAGCATCCCCAAGCAATGAAGCATTTGCAAAAGCA 1096
 QY 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
 DB 1097 AGAGGAGCAGCTTATGAATCTTCAAGATATTTCACAATAAACCACCAAGCATTGACAGTAT 1156
 QY 381 SerLysSerGlyHisLysLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
 DB 1157 TCGAAGAGTGGACATAAACCCAGATAATATTAAAGGAAATTTGGAATTCAAAAATGTTTCA 1216
 QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
 DB 1217 TTCAGTTACCCCTTCGAAAAGAAAGTTAAAGATCTTAAAGGCTCTCAACCTGAAGTTTCAG 1276
 QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
 DB 1277 AGTGGCAGACAGTGGCGCTGGTTGGGAACAGTGGCTGCGGGAAGACACAGCCGTGAC 1336
 QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
 DB 1337 CTGATGCAGAGGCTCTATGACCCACAGATGGCATGCTGTATGATGACAGGACATT 1396
 QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
 DB 1397 AGGACCATTAATGTPAGGCATCTTCGGGAAATTTACTGGTGGTGGTCTAGGAGCCCTGTG 1456
 QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
 DB 1457 TTGTTTCCACCACGATAGCTGAAAACATTCGCTATGGCCGCGAAGAAATGTCACCATGGAT 1516
 QY 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
 DB 1517 -GAGATTGAGAAAGCTGTTAAGGAAGCCATGCCTATGATTTTATCATGAACTACCTAAT 1576

Db 3737 AATGCAGATTAAATAGTGGTGTTCAGAAATGCAAGTCAAGGAGCATGGCACACATCAA 3736
 QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
 Db 3797 CAGCTGCTGGCCAGAAAGGCATCTATTTTCCATGTCAGTGTCCAGGCTGGAGCAAG 3856
 QY 1281 Arg 1281
 Db 3857 CGC 3859
 RESULT 5
 AAD03489
 ID AAD03489 standard; cDNA; 4317 BP.
 AC AAD03489;
 DT 13-JUN-2001 (first entry)
 DE Dog P-glycoprotein (PGP) cDNA #2.
 KW Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
 KW drug bioavailability; transgenic animal; genetic model; ss.
 OS Canis familiaris.
 XX
 FH Key Location/Qualifiers
 FT CDS 70..3912
 FT /*tag= a
 FT /product= "Dog P-glycoprotein (PGP) #2"
 XX
 PN WO200123540-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26767.
 XX
 PR 28-SEP-1999; 99US-0156510.
 XX
 PA (GENT-) GENTEST CORP.
 XX
 PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 XX
 DR WPI; 2001-235373/24.
 DR P-PSDB; AAE00304.
 XX
 PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors -
 XX
 PS Claim 1; Page 66-72; 11lpp; English.
 XX
 CC The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (PGP) cDNA. The
 CC PGP enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.
 XX
 SQ Sequence 4317 BP; 1293 A; 844 C; 1019 G; 1161 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 4317

Score: 6425.50 Matches: 1273
 Percent Similarity: 99.69% Conservative: 4
 Best Local Similarity: 99.38% Mismatches: 3
 Query Match: 99.25% Indels: 1
 DB: 22 Gaps: 1
 US-09-672-725C-23 (1-1281) x AAD03489 (1-4317)
 QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
 Db 70 ATGGATCCTGAAGAGGCCGTAAGGGAGTGCAGAGAAGAACTTCTGGAAAATGGGCAAA 129
 QY 21 LysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
 Db 130 AAAAGTAAAAA--CAGAGAAAGAAAGAAACCAACTGTCCAGCAGCTTGCAATGTTT 186
 QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLallelle 60
 Db 187 CGCTATTCAAATTTGGCTTGATAGTTGTATATGTTGGTGGGACAAATGGCTGCCATCATC 246
 QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
 Db 247 CATGGAGCTGCACCTCCCTCTCATGATGCTGGTGTGTTTGGAAACATGACATAGCTTTGCA 306
 QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
 Db 307 AATGCAGGAATTTCAAGAAACAAACCTTTCCAGTTATTAATTAAGAAAGTATACGAAC 366
 QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120
 Db 367 AATACACAACATTTTCATCAACCATCTGGAGAGGAAATGACCACGCTATGCCCTATTATTAC 426
 QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
 Db 427 AGTGGGATCGGTGCTGGCTGGTGGCTGCTTACATCCAGTTTCATCTGGTGGCTTG 486
 QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
 Db 487 GCAGCAGGAAGACAGATACCTCAAAATTAAGAAACAAATTTTTCATGCTATCATGGACAG 546
 QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
 Db 547 GAGATTGGCTGGTTCACGTGCATGACGTGGGAGCTTAACACCGCGCTCACAGACGAT 606
 QY 181 ValSerLysLysLeuGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
 Db 607 GTCCTCAAAATCAATGAAGAAATGGCGACAAGTTGGAGTGTCTTTCAATCAATAGCA 666
 QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
 Db 667 ACATTTTTCACCGGTTTATAGTGGGTTTACACCTGGTGGAGCTAACCTTGTGATT 726
 QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
 Db 727 TTGGCCATCAGCCCTGCTTCTGGACTTTTCAGCCGCATCTGGGCAAGATATCTCTTCA 786
 QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaIleGluValLeu 260
 Db 787 TTTACTGATAAAGAACTCTTGCCCTATGCAAAAGCTGGAGCAGTAGCTGAAGATCTTA 846
 QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
 Db 847 GCAGCAATCAGAACTGTGATTGCCCTTGAGGACAAAGAAAGAACTTGAAGGTACAAAC 906
 QY 281 LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
 Db 907 AAAAAATTTAGAGAAGCTAAAAAGAAATGGGATAAAGAAAGCTATCACGCCCAACATTCT 966
 QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
 Db 967 ATGTGTGGCGCTTCTTCTTATGATCATCATATGCTATGCTTGGCTTCTGTATGGGACC 1026
 QY 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340

Db 1027 TCCCTGGTCCCTCCAGTGAATATACTATTGGACAGGTACTCACTGCTCTCTTCTGTGTA 1086
QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db 1087 TTAATTGGGGCTTTTAGTATTGGACAGCATCCCCCAAGCAATTGAAGCAATTTGCAACGCA 1146
QY 361 ArgGlyAlaAlaTyrGluIlePheIleYsIleIleAspAsnLysProSerIleAspSerTyr 380
Db 1147 AGAGAGCAGCTTATGAATCTTCAAGATAATTGACAATAAACCAAGCAATTCAGACGCTAT 1206
QY 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1207 TCGAAGAGTGGACATAAACCAAGATTAATTAAAGGAAATTTGGAATTCAAAAATGTTTAC 1266
QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1267 TTCAGTTACCTTCTCGAAAGAAAGTAAAGATCTTAAAGGGTCTCAACCTGAAGGTTTACG 1326
QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1327 AGTGGCAGACAGTGGCGTGTGGGAACAGTGGCTGGGGAAGAGCACCACCGTGCAG 1386
QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db 1387 CTGATCCAGAGGCTCTATGACCCACAGATGCGATGGTCTGTATTGATGGACAGGACATT 1446
QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db 1447 AGGACCATAAATGTAAGGCATCTTCGGGAAATTTACTGTGTGGTGCAGTCAGAGCCTGTG 1506
QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1507 TTGTTTCCACCCAGATAGCTGAAACATCTCCTATGCGCCGCGAATAATGTCACCATGAT 1566
QY 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1567 GAGATTGAGAAAGCTTTAAGGAAGCAATGCTATGATTTTATCATGAAACTACCTAAT 1626
QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
Db 1627 AAATTTGACACACTGGTGTGGAGAGAGGGGGCCCGGCTGAGTGGGACAGAAACAGAGA 1686
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
Db 1687 ATCGCCATTGCTCGGGCCCTGGTTCGCAACCCCAAGATTCTTCTGCTGGATGAGCAACG 1746
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1747 TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCTGGATAAGGCCAGAAAA 1806
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1807 GGGCGGACTACCACTGTGATAGCTATCTGTTGTCTACAGTTCGTAATGCCGATGCTATT 1866
QY 601 AlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1867 GCTGGTTTTGATGATGGAGCTATTGTGGAGAAAGGAAATCATGATCAACTCATGAAGAG 1926
QY 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1927 AAGGGCATTTACTTCAAACTTGTCAATTCAGACAAGAGGAAATGAAATTCAGTTAGAA 1986
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1987 AATGCCACTGGGTGAATCCAAAGTGAAGTGAATGATGCTTGGAAATGTCCTCAAAAGATTCA 2046
QY 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db 2047 GGGTCCAGTTTAATAAAAGAGATCAACTCGCAGGAGTATACATGTCACCAAGGCCAA 2106
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTyr 700

Db 2107 GACAGAAAGCTTGGTACAAAAGAGAGACTTGAATGAGAAATGTTCCCTTCAGTTCTTCTGTG 2166
QY 701 ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 720
Db 2167 AGGATTCGAAGCTGAACCTCAACTGAATGGCCTTATTTTGTGGTTGGTATATTTTGTGCT 2226
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2227 ATTATAAACCGGAGCCCTGCAACCCAGCATTTTCAATAATATTTTCAAGGATTATAGGATC 2286
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2287 TTTACCCGAGATAGGATCCTGAAACAAAACAGACAATAGTAACATGTTTCTGTATTG 2346
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2347 TTTCTAGTCTCTGGAATTTATTTCTTTTATTACATTTTCTCTCCAGGGCTTCACATTGGC 2406
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2407 AAAGCTGGGAGATCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTGAGA 2466
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2467 CAGGATGTCAGCTGGTGTGATGACCCCTAAACACACCACCTGGAGCATGACACCGGCTT 2526
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2527 GCCAATGATCGGCTCAAGTTAAAGGGCTATAGGTTCAGGCTTGTGTCATTACCCAG 2586
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTyrGlnLeuThr 860
Db 2587 AATATAGCAATCTCGGACAGGCATTATTATATCTTAATCTATGTTGGCAATTAAACA 2646
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2647 CTTTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGAGTGTGTAATGAAATG 2706
QY 881 LeuSerGlyGlnAlaLeuLysAspLysGluLeuGlyAlaGlyLysIleAlaThr 900
Db 2707 TTGCTCGACAGACACTGAAAGATAAGAAAGAGCTAAGAGAGCTGGGAAGATTGTCTACA 2766
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2767 GAAGCCATCGAAAACCTTCGCAACTGTTGTTCTTGTACTCGGAGCAGAGTTTGAATAC 2826
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2827 ATGATGCACAGAGTTTGCAGTACCATACAGAACTCTTTGAGGAAAGCACACATCTTC 2886
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2887 GGGGTCTCATTTCTATCAACCCAGGCAATGATGATTTTTCCTATGCTGGCTGTTCGG 2946
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 980
Db 2947 TTTGGTGCCTACTTGGTGGCAAAATGAGTTTCATGAACCTTCAGGATGTTCTTTTGGTATTC 3006
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
Db 3007 TCAGCTATTGCTTTGGTGCCATGCGATGGCGAGCTCAGTTCATTGCTCTCGACTAT 3066
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
Db 3067 GCCAAAGCAAGATATCAGCAGCCAGCTCATCATGATCATTTGAAAAAGCCCTCTGATT 3126
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3127 GACAGCTACAGCCCTCAGCGCCTCAAGCCAAATACGTTGGAAGGAATGTGACATTTAAT 3186
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3187 GAGTCTGGTTCAACTATATCCACATCGACACAGATCCCGTGTCCAGGGGCTGAGCCTC 3246

Db 2332 GGCAAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATGTTTCCGATCCATGCTC 2391
Qy 800 ArgGlnAspValSerTyrPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2392 AGACAGGATGTGAGTTGGTTTGATGACCCCTAAACACCACTGGAGCATTTGACTACAGG 2451
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2452 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTGCTGTAAATTACC 2511
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleThrGlyTrpGlnLeu 859
Db 2512 CAGAAATAGCAATCTTGGACAGGAATAATTATATCTTCTATCTATGCTGGCAACTA 2571
Qy 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2572 ACATGTTTACTCTAGCAATGTGACCATCATTTGCATATACAGAGTTGTTGAATGA 2631
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2632 ATGTTGTCTGGACAAGCACCTGAAGATAAGAAAGACTAGAAAGTCTGGGAAGATCGCT 2691
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2692 ACTGAAGCAATAGAAACTTCCGAACCGTTGTTCTTGTGACTCAGGAGCAGAAAGTTTGA 2751
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2752 CATATGTATGCTCAGAGTTTGCAGGTACCATACAGAACTCTTTGAGGAAGCACACATC 2811
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2812 TTTTGGAAATTACATTTTCTTCCATCCAGGCAATGATGATTTTCTTATGCTGGATGTTT 2871
Qy 960 ArgPheGlyValTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 2872 CGGTTTGGAGCCCTACTTGTGGGCACATAAACTCATCAGCTTTGAGGATTTCTGTAGTA 2931
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 2932 TTTTTCAGCTGTGCTTTGTCGTCATGGCGGTGGGCAAGTCACTGTCATTTGCTCTGAC 2991
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 2992 TATGCCAAAGCCAAATATACAGCAGCCACATCATCATCATGATGATGATAAAACCCCTTG 3051
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3052 ATTGACAGCTACAGCAGGAAGCCCTAATGCCGACACACATTTGGAAGGAATGTCATTT 3111
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3112 GGTGAAGTTGTATTCAACTATCCACCCGACGGACATCCAGTGTTCAGGAGCTGAGC 3171
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3172 CTGGAGGTGAAGAGGGCCAGACGCTGCTGCTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3231
Qy 1080 ThrValValGlnLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3232 ACAGTGGTCCACTCTCTGGAGCGGTCTACGACCCCTTGGCAGGGAAGTCTGCTTGTAT 3291
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3292 GGCAAAGAAATAAGCGACTGAATCTCAGTGGCTCCGAGCACACCTGGGCATCGTGTC 3351
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3352 CAGGAGCCCATCTCTTTGACTGACGATGCTGAGAACATTTGCTATGAGACAACAGC 3411
Qy 1140 ArgValValSerHisGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
Db 3412 CGGGTGGTGTACAGAGAGAGTCTGAGGGCAGCAAGGAGGCAACATACATACATGCCTTC 3471

QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3472 ATCGAGTCACTGCCTTAATAATATAGCACTAAAGTAGGAGACAAGAACTCAGCTCTCT 3531
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3532 GGTGGCCAGAAACACGCAATTGCCATAGCTCGTGCCTTGTGTAGACAGCTCATATTTTG 3591
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3592 CTTTTCGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGGTTGTCCAGAGCC 3651
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3652 CTGGCAAAAGCCAGAGAGGCGCCACCTGTCATTTGATTTGCTCAGCGCTGTCCACCATC 3711
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3712 CAGAATGCAGACTTAATAGTGTGTTTCAGATGCGCAGTCAAGAGCATGGCAGCAT 3771
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3772 CAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGGTTCAGTGTCAGGCTGGAACA 3831
QY 1280 LysArg 1281
Db 3832 AAGCGC 3837
RESULT 7
ABA94365
ID ABA94365 standard; DNA; 3860 BP.
XX ABA94365;
AC ABA94365;
DT 26-MAR-2002 (first entry)
XX Human BCRP DNA related seq Id No. 1.
DE Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
KW hepatotropic; neurodegenerative; cytosolic; antianemic; muscular; BCRP;
KW cardiant; gene therapy; ds.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..3843
FT /*tag= a
XX
PN WO200192877-A2.
XX
PD 06-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US17459.
XX
PR 31-MAY-2000; 2000US-0584586.
PR 29-MAY-2001; 2001US-0866866.
XX
XX (SUJD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX Sorrentino B, Schuetz J;
XX WPI: 2002-114368/15.
XX P-PSDB; ABB07266.
XX
PT Identifying a stem cell, for treating e.g., muscular dystrophy,
PT myocardial infarction, Parkinson's disease, or neurodegenerative
PT disorders, comprises detecting the expression of an ATP transport
PT protein (BCRP) by a cell
XX
PS Disclosure; Page 53-55; 87pp; English.
XX
CC The invention provides a method of identifying and/or isolating a stem

Db 1792 ATCCGCTGGTTTCGATGATGAGTCAATGTGGAGAAAGAAATCATGATGAATCAATGAA 1851
Qy GluLysGlyIleTyrPheLysLeuValThrMetClnThrArgGlyAsnGluIleGluLeu 639
Db 1852 GAGAAAGCATTTACTTCAAACTTGTCCAAATGTCAGACAGCAGAAATGAAGTTGAATTA 1911
Qy GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 1912 GAAATCAGCTGATGAATCCAAAGTGAATGATGCCCTTGAAATGCTTCAATGAT 1971
Qy SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 1972 TCAGATCCAGTCTAATAAGAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAGCC 2031
Qy GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2032 CAAGACAGAAGCTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCCTT 2091
Qy TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2092 TGGAGGATTATGAAGCTAAATTTAACTGAATGGCTTATTTTCTGTGGTGTATTTGT 2151
Qy AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2152 GCCATTATAATGGAGGCTTGCACACAGCATTTGCAATAATATTTTCAAGATTATAGG 2211
Qy IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2212 GTTTTACAGAATTGATGATCTCTGAACAAACGACAGCAAGTAGTAACTGTGTTTCACTA 2271
Qy LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2272 TTGTTTCTAGCCCTTGAATTAATTTCTTTTATATACATTTTCTTCAAGGTTTTCACATTT 2331
Qy GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2332 GGCAGAGCTGGAGAGATCTCCACAGCGCTCCGATACATGATGTTTCCGATCCATGCTC 2391
Qy ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2392 AGACAGATGTGAGTTGGTTGATGACCTTAAACACACACCTGGACATGACTACCAGG 2451
Qy LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2452 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTCTGCTAATATAC 2511
Qy GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2512 CAGAAATATAGCAATCTCGGACAGGAATATATATCTTCATCTATGTTGGCAACTTA 2571
Qy ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2572 ACACGTGTTACTCTAGCAATTTACCCATCTGCAATAGCAGGAGTTGTTGAATGAA 2631
Qy MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2632 AGTTGCTGGACAGCACTGAAAGATAGAAGAACTAGAAAGTGCTGGGAAGATCGCT 2691
Qy ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2692 ACTGAAGCAATAGAAAACCTTCGACCGTTGTTCTTTTGACTCAGACGACAGAGTTTGA 2751
Qy TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2752 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAACTCTTTGAGGAAAGCACACATC 2811
Qy PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2812 TTTGGAAATACATTTTCTTCCACGCAAGATGATGATTTTCTATGCTGGATGTTTC 2871
Qy ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979

Db 2872 CGGTTTGGAGCCCTACTTGTGGTCACATAAATCATGACGCTTTGAGATGTTCTGTAGTA 2931
Qy PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 2932 TTTTCAGCTGTTGTCTTGTGGTCCATGGCGTGGGCAAGTCAGTTCAATTTGCTCCTGAC 2991
Qy TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 2992 TATGCCAAAGCCAAATATACAGACCCCATCATCATGATCATTTGAAAAAAACCCCTTG 3051
Qy IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3052 ATTGACAGCTACAGCAGGAGGCCCTAATGCGCAACACATTTGGAAGAAATGTCACATTT 3111
Qy AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3112 GGTGAAGTTGTATTCAACTATCCACCGACCGACATCCCATGCTCTTCAGGAGTACG 3171
Qy LeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3172 CTGGAGGTGAAGAGGGCCAGACGCTGGCTCTGGTGGCAGCAGTGCGTGGGAAGAGC 3231
Qy ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3232 ACAGTGGTCCAGCTCTCGAGCGGTTCTACGACCCCTTGGCAGGGAAGTGTGCTTGTAT 3291
Qy GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3292 GGCAGGAAATTAAGCGACTGAATGTTCAAGTGGCTCCAGGACACACCTGGGGATCGTCTCC 3351
Qy GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3352 CAGGAGCCCATCTGTTTGACTGACGACATGCTGAGAACATGCTGATGGAGACACAGC 3411
Qy ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisPhe 1159
Db 3412 CGGTTGGTGTACAGAAAGAGATCGTAGGGCAGCAAGGAGGCAACATACATGCTCTCT 3471
Qy IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3472 ATCGAGTCACTGCTTAATATATAGCACTAAGTAGGAGACAAAGGAACCTCAGCTCTCT 3531
Qy GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3532 GTGGCCCAAGAAACAACGATTCCTAGCTGTCCTGCTGTTAGACAGCCTCATATTTTG 3591
Qy LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3592 CTTTGGATGAAGCCACGCTCAGCTCTGGATACAGAAAGTGAAGGTTGTCCAAGAGCC 3651
Qy LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3652 CTGACAAAGCCACAGAGGCGGACCTGATGATGCTGCTACCCGCTGTCCACCATC 3711
Qy GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGlyThrHis 1259
Db 3712 CAGAATGACAGCTTAATAGTGTGCTTTCAGAAATGGCAGAGTCAAGGAGCATGGCAGCAT 3771
Qy GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3772 CAGCAGCTGTGGCAGCAAGAGGATCTATTTTTTCAATGTGCTGCTCCAGGCTGGAACA 3831
Qy LysArg 1281
Db 3832 AAGCGC 3837
RESULT 8
AAF86127
ID AAF86127 standard; cdna; 4186 BP.
XX
AC AAF86127;
XX
DT 25-JUN-2001 (first entry)

QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrVal 439
Db 1351 CAGAGTGGCAGAGGGTGGCCCTGGTTGGAAACAGCGGCTGTGGGAAGACCAACGGTC 1410
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1411 CAGCTGATGACAGAGGCTTTATGACCCACACAGAGGGCATGTCAGTGTGTGACAGGAT 1470
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1471 ATTAGGACCAATAACGTAAAGGTTTACGGGAAATCATCGGTGTGTGATCAGGAACCT 1530
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1531 GTATTGTTGCCACCAGATAGCTGAAACATTCGCTATGCTGCTGAAGATGTCACCATG 1590
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1591 GATGAGATTGAGAAAGCTGCAAGGAAGCCAATGCCATGCTTATCATGAAACTGCCT 1650
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlnLysGln 539
Db 1651 CAGAAATTGACACCTGGTTGAGAGAGAGGGGCCAGCTGAGTGGTGGGAGAACGAG 1710
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 1711 AGGATGCCATTGACGTCGCCCTGGTTGCCAACCCCAAGATCCTCCTGCTGGACGAGGCC 1770
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 1771 ACGTGAGCCTTGACACAGAAAGTGAAGCAGTGGTTGAGTGGCTCTGGATAAGGCCAGA 1830
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 1831 AAAGGTCGACACCACTTGTAGTCTCATCTGTTGCTACCGTTCGTAAATGCCGACGTC 1890
QY 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 1891 ATCGCTGGTTCGATGATGAGTCAATTGTGGAGAAAGAAATCATGATGAGCTCATGAAA 1950
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 1951 GAGAAAGGCATTACTTCAAACTGTCAATGTCAGACAGCAGGAAATGAAATTA 2010
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2011 GAAAAATGACGCTGTAATCAAAAGTGAATGATACCTTGGAAATGCTTCACATGAT 2070
QY 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2071 TCAGGATCCAGCTGTAATAGAAAAAGATCCACTCGTAGGATGTCCTGGATCAACAGGC 2130
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2131 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 2190
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2191 TGGAGGATTATGACGTAATTAATCACTGAGTGGCCTTATTTGTTGTTGGTATTTTGT 2250
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2251 GCCATTATAAATGGAGGCTGCAACAGCAATTCAGTAATATTTTCAAGATTATAGG 2310
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2311 ATTTTACAGAAATGATGATGCCGAAACAAACAGACAGATAGTAACCTGTTTTCACATA 2370
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2371 TTGTTTCTAGTCTTGGAAATGTTCTTTATATACATTTTTCCTTCAGGGCTTCACATTT 2430
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799

Db 2431 GGCAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATGGTTTCCGATCCATGCTC 2490
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2491 AGACAGGATGTGAGCTGGTTGATGACCTAAAAACACCACCTGGAGCATGTACTACCAGG 2550
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2551 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGTTCACAGCTTCGATATAATTACC 2610
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2611 CAGAAATATAGCAAAATCTTGGGACAGAAATATATATCTTAATCTATGGTTGGCAACTG 2670
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2671 ACACGTGTACTCTTAGCAATTTGACCCATCATGCAATAGCAGAGATTGTTGAATGAAA 2730
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2731 ATGTTGCTCTGGACAAAGCACTGAAAGATAGAAAGAACTAGAAAGTCTGGGAAGATCGCT 2790
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2791 ACTGAAGCAATAGAAAATCTCCGAACTGTTGTTCTTTGACTCAGGAGCAGAACTTTGAA 2850
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2851 CATATGTTATGATCAGAGTTTCGAGGTACCATACAGAACTCTTTGAGGAAAGACACATC 2910
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2911 TTTGGAATCACGCTTTCCTCCACGCAAGCAATGATGATTTTCTCTATGCTGGATGTTTC 2970
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 2971 CGEITTTGGAGCCTACTTGTGGCACATAGTCTCATGAGCTTTGAGGATGTTCTGTTAGTA 3030
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3031 TTTTCAGCTGTGTTCTTTGGTGCCATGCGCTGGGCAAGTCAGTTTCATTTGCTCCTGAC 3090
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3091 TATGCCAAAGCCAAAGTATCAGACGCCACATCATCATGATCATTTGAAAAAACCCTTTG 3150
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3151 ATTCACAGCTACACACAGAGGCTTAAGCCGCAACACATTTGGAAAGAAATGTCACATTT 3210
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3211 AATGAAGTTGTATTCAACTATCCACCCGATCGACATCCAGTGTCTCAGGGGCTGAGC 3270
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3271 CTGGAAGTGAAGAAAGGCCAGACGCTGGCCCTGGTGGGCGACAGTGGCTGTGGGAAGAGC 3330
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3331 ACGGTGCTCCAGCTCCTGGAGCGGTTCTATGACCCCTTGGCGGGGAAAGTCTGCTTGAC 3390
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3391 GGCAAGAAATAAAACAACTGAATGTTCACTGGCTCCGAGCACACCTGGGCACTCGGTGCC 3450
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3451 CAGAGGCCCATCTGTTGACTGCAGCATTTAGTGAACAATTCCTATGCCCTATGGAGCAACAGC 3510
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159

Db 3511 CGGGTGGTGCACAGGAAGACATCGTAGGGCCAGCCAGGAGGCCAATATACAGCCCTC 3570

Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179

Db 3571 ATCGAGTCACCTGCTTAATAATATAGCACACAGAGTAGGAGACAAAGAACTTCAGCTCTCT 3630

Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199

Db 3631 GGTGCCAGAAACACGCGATTGCCATAGCTCGTGCCCTTGTAGACAGCCATATATTG 3690

Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219

Db 3691 CTTTGGATGAGCCACATCAGCTCTGCATACAGAAAGTGAAGGTTGTCCAAAGAGCC 3750

Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239

Db 3751 CTTGCAAGCCAGAGAGCGGTACCTGCATTGTGTATGTCTCCCGCTCTCCACCATC 3810

Qy 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259

Db 3811 CAGATGCAGACTTATAGTGTGTTTCAGATGGCAGAGTCAAGGACGACGGCACACAT 3870

Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279

Db 3871 CAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAGCA 3930

Qy 1280 LysArg 1281

Db 3931 AAQCGC 3936

RESULT 9

AAF86128

ID AAF86128 standard; cDNA; 4195 BP.

XX

AC AAF86128;

DT 25-JUN-2001 (first entry)

XX

DE Cynomologous monkey P-glycoprotein cDNA variant 1.

XX

KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1; efflux pump; ss.

KW

OS Macaca fascicularis.

FH

FT Key Location/Qualifiers

FT CDS 100..3951

FT /*tag= a "pgp"

FT /product= "pgp"

FT /note= "P-glycoprotein"

FT misc_feature 376..384

FT /*tag= b

FT /note= "Insertion of 9 nucleotides relative to PGP allelic variant AAF86127"

XX

PN W0200123565-A1.

XX

XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-US26592.

XX

XX

PR 28-SEP-1999; 99US-0156921.

PR

PR 12-OCT-1999; 99US-0158818.

XX

XX

PA (GENT-) GENTEST CORP.

XX

PI Stocker PJ, Steimel-Crespi DT, Crespi CJ;

XX

XX

DR WP1; 2001-316136/33.

DR

DR P-PSDB; AAB81065.

XX

XX

PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein (PGP) and homologous PGP polypeptides are useful for predicting

PT bioavailability of compound and increasing PGP transporter activity in cell

XX

XX Example 1; Page 59-65; 84pp; English.

XX

CC This invention relates to a polynucleotide sequence encoding a cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the PGP protein. PGP, also known as multidrug transporter, MDR1 is a member of the ABC transporter superfamily. The enzyme serves as an efflux pump exporting small molecules across the cell membrane. The invention includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence and protein, and also that of an allelic variant. The PGP polynucleotide sequence is useful for increasing PGP transporter activity in a cell. Antisense sequences of the cDNA are useful for inhibiting PGP transport activity in a mammalian cell. They may also be used for increasing the bioavailability of a drug. The present sequence represents cDNA encoding cynomologous monkey P-glycoprotein. This sequence contains a 9 nucleotide insert compared to the PGP allelic variant given in AAF86127.

XX

SQ Sequence 4195 BP; 1230 A; 801 C; 1039 G; 1125 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 4195

Score: 5859.50 Matches: 1165

Percent Similarity: 94.80% Conservative: 57

Best Local Similarity: 90.38% Mismatches: 52

Query Match: 90.51% Indels: 15

DB: 22 Gaps: 5

US-09-672-725c-23 (1-1281) x AAF86128 (1-4195)

Qy 1 MetAspProGluGlyArgLysGlySerAlaGlu---LysAsnPheTyrPLeuMetGly 19

Db 100 ATGGATCTTGAAGGGGACCCCAATGGAGGAGAGAGAAAGAACTTTTAACTGAAC 159

Qy 20 LysLysSerLysLysAsnGluLysLysLysLysProThrValSerThrPheAlaMet 39

Db 160 AATAAAGTAAAAA---GATAAGAGGAAGAAACCAACTGTCAGTGTATTTCATG 216

Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59

Db 217 TTTCGCTATTCAAAATGGCTTGACAAGTTGTATATGGTGGGAACTTTGGTGCATC 276

Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79

Db 277 ATCCATGGAGCTGCAGCTTCCTCTCATGATGCTGCTGTTGGAGACATGACGGATACCTT 336

Qy 80 AlaAsnAlaGly-----IleSerArgAsnLysThrPheProVal 92

Db 337 GCAATGCGAGAAATTTAGGAGATTTAGGAGCTCTGTTGTTTAAACAACACT----- 387

Qy 93 IleIleAsnGluSerIleThrAsnAsnThrGlnHisPheIleAsnHisLeuGluGlu 112

Db 388 -----AATAGCAGTAATATCATCTGATACATGCTCCGCTCATGAAT---CTGGAGCAGAT 438

Qy 113 MetThrThrTyrAlaTyrTyrTyrSerGlyIleGlyAlaGlyValLeuValAlaIleTyr 132

Db 439 ATGACCGATGATGCTTATATACAGTGAATTTGCTGGGCTGCTGCTGCTGCTTAC 498

Qy 133 IleGlnValSerPheTrpCysLeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGln 152

Db 499 ATTCAGTTTTCATTTTGGTGGCTGGCAGTGAAGACAAATAGAAAAACAG 558

Qy 153 PhePheHisAlaIleMetArgGlnIleGlyTyrPheAspValHisAspValGlyGlu 172

Db 559 TTTTTCATGCTATTAATGCGACGAGATAGCTGTTTGTATGTCAGAGATGTTGGGAG 618

Qy 173 LeuAsnThrArgLeuThrAspValSerLysIleAsnGluGlyIleGlyAspLysIle 192

Db 619 CTTAACCCGCGCTTACAGATGATGCTCCAAGATTAATGAAGAAATTTGGTGCACAAAAT 678

Qy 193 GlyMetPhePheGlnSerIleAlaThrPhePheThrGlyPheIleValGlyPheThrArg 212

Db 679 GGAATGTTCTTTCAGTCAATGGCAACATTTTTCACCTGGGTTTATAGTAGGATTTACACGT 738
Qy 213 GlyTrpLysLeuThrLeuValIleLeuAlaIleSerProValLeuGlyLeuSerAlaAla 232
Db 739 GGTGGAGGTACCCCTGTGTGATTTGGCCATCAGTCCCTGTTCTTGACTGTGCAGTGA 798
Qy 233 IleTrpAlaLysIleLeuSerSerPheThrAspLysGluLeuLeuAlaTyAlaLysAla 252
Db 799 GTCTGGCAAGATACTGTCTTCATTTACTGATAAGAACTCTAGCTATGCAAAAGCT 858
Qy 253 GlyAlaValAlaGluGluValLeuAlaAlaIleArgThrValIleAlaPheGlyGlyGln 272
Db 859 GGAGCAGTAGCTGAAGAGGCTTTGGCAGCAATTAGAACTGTGATTGCATTTGGAGACAA 918
Qy 273 LysLysGluLeuGluArgTyAsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLys 292
Db 919 AAGAAAGACTCGAAGGTACACAAAATTTAGAAGAAGCTTAAAGAATTTGGGATAAG 978
Qy 293 LysAlaIleThrAlaAsnIleSerIleGlyAlaAlaPheLeuLeuIleTyAlaSerTy 312
Db 979 AAAGCTATTACAGCCAATATTTCTATAGTGTGCTGCTTCTCTGCTTATCTATCATCTTAT 1038
Qy 313 AlaLeuAlaPheTrpTyGlyThrSerLeuValLeuSerSerGluTySerIleGlyGln 332
Db 1039 GCTCTGGCCTTCGGTATGGGACCACTTGGTCTCTCAAGGAATATTCTATTGGACAA 1098
Qy 333 ValLeuThrValPhePheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerPro 352
Db 1099 GTACTCACTGTATCTTTCTGTATTAAATGGGCTTTTAGTGTGGACAGCATCTCCA 1158
Qy 353 SerIleGluAlaPheAlaAsnAlaArgGlyAlaAlaTyGlyIlePheLysIleIleAsp 372
Db 1159 AGCATTGAAGCATTTGCAAAATCAAGAGCAGCAGCTTTTGAAATCTCAAGATAATTGAT 1218
Qy 373 AsnLysProSerIleAspSerTySerLysSerGlyHisLysProAspAsnIleLysGly 392
Db 1219 AATAAGCCAGATTTGACAGCTATTGGAAGATGGGCAACAAACAGATAATATTAAAGGA 1278
Qy 393 AsnLeuGluPheLysAsnValHisPheSerTyProSerArgLysGluValLysIleLeu 412
Db 1279 AATTTGGAATTCAGAAATGTTCACTTCAGTTACCCATCTCGAAAGAGAGTTAAGATCTTG 1338
Qy 413 LysGlyLeuAsnLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGly 432
Db 1339 AAGGCGCTGAACCTGAAGGTGCAGAGTGGCAGACGGTGGCCCTGTTGGAAACAGCGCG 1398
Qy 433 CysGlyLysSerThrThrValGlnLeuMetGlnArgLeuTyArgLeuTyArgProThrAspGlyMet 452
Db 1399 TGTGGAGAGACCAACCGTCCAGCTGATGACAGGCTTTATGACCCACACAGAGGGCATG 1458
Qy 453 ValCysIleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThr 472
Db 1459 GTCAGTGTGTGAGACAGATATTAGGACCATTAACGTTAAGGTTCTACGGGAATCATC 1518
Qy 473 GlyValValSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTy 492
Db 1519 GGTGTGGTGGAGTCAAGAACCTGTATTGTTGCCACGATAGCTGAAACATTCGCTAT 1578
Qy 493 GlyArgGluAsnValThrMetAspGluIleGluLysAlaValLysGluAlaAsnAlaTy 512
Db 1579 GGTCTGAAGATGTCACCATGGATGAGATTGAGAAGCTGTCAAGGAAGCCCAATGCCTAT 1638
Qy 513 AspPheIleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGln 532
Db 1639 GACTTTATCATGAAGTTCAGAAATTTGACACCCCTGGTTGGAGAGAGAGGGGCCAG 1698
Qy 533 LeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLys 552
Db 1699 CTGAGTGTGGCAGAGACAGAGATGCCATTGCAGTGCCTTGGTTCGCAACCCCAAG 1758
Qy 553 IleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGln 572
Db 1759 ATCCTCTGCTGGAGGAGGCCCGTCAGCCTTGGACAGAAAGTAGGAGCGTGGTTTCAG 1818

Qy 573 ValAlaLeuAspLysAlaAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSer 592
Db 1819 GTGGTCTCTGATAAGGCCAGAAAAGTCCGACCACCATTTGTGATAGCTCATCGTTTGCT 1878
Qy 593 ThrValArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValIleGlyGly 612
Db 1879 ACGGTTCGTAATGGCAGCTCATCGTGGTTCGATGATGGAGTCAATTTGTGGAGAAAGGA 1938
Qy 613 AsnHisAspGluLeuMetLysGluLysGlyIleTyPheLysLeuValThrMetGlnThr 632
Db 1939 AATCATGATGAGCTCATGAAAGAGAAAGCACTTACTTCAAACTGTGCATGCGAGACA 1998
Qy 633 ArgGlyAsnGluIleGluLeuAsnAlaThrGlyGluSerLysSerGluSerAspAla 652
Db 1999 GCAGAAATGAAATTTGAATTTAGAAATGCAGTGAATCCAAAAGTGAATGATACC 2058
Qy 653 LeuGluMetSerProLysAspSerGlySerSerLeuIleLysArgArgSerThrArgArg 672
Db 2059 TTGGAAATGCTTTCACATGATTCAGATCCAGTCTTAATAAGAAAAGATCCACCTCGTAGG 2118
Qy 673 SerIleHisAlaProGlnGlyGlnAspArgLysLeuGlyThrLysGluAspLeuAsnGlu 692
Db 2119 AGTGTCCGTGGATCACAAAGCCACAGACAGAAAGCTTAGTACCAGAGGCTCTGGATGAA 2178
Qy 693 AsnValProProValSerPheTrpArgIleLeuLysLeuAsnSerThrGluTrpProTy 712
Db 2179 AGTATACCTCCAGTTCTCCTTTTGGAGGATTAAGAAGTAAATTTAACTAGTGGCTTAT 2238
Qy 713 PheValValGlyIlePheCysAlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIle 732
Db 2239 TTTGTTGTTGGTGATTTTGTGCCATTATAATGGAGTCTGCAACAGCATTTGCGAGTA 2298
Qy 733 IlePheSerArgIleIleGlyIlePheThrArgAspGluAspProGluThrLysArgGln 752
Db 2299 ATATTTTCAAGATTAATAGGATTTTACAAGAAATGATGATGATGATGATGATGATGATGAT 2358
Qy 753 AsnSerAsnMetPheSerValLeuPheLeuValLeuGlyIleIleSerPheIleThrPhe 772
Db 2359 AATAGTAACCTTTGTTTCTCACTATTTCTTCTAGTCTCTGGAATTTGTTCTTTTATACATTT 2418
Qy 773 PheLeuGlnGlyPheThrPheGlyLysAlaGlyGluIleLeuThrLysArgLeuArgTy 792
Db 2419 TTCCTTCAGGCTTCACATTTGGCAAGCTGGAGAGATCTCCACCAAGCGCTCCGATAC 2478
Qy 793 MetValPheArgSerMetLeuArgGlnAspValSerTrpPheAspAspProLysAsnThr 812
Db 2479 ATGTTTTCCGATCCATGCTCAGACAGGATGTGAGCTGTTTGTATGATGATGATGATGATGAT 2538
Qy 813 ThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaGlnValLysGlyAlaIleGly 832
Db 2539 ACTGGAGCATTGACTACAGGCTCGCCAAATGATGCTGCTCAAGTTAAAGGGGCTATAGT 2598
Qy 833 SerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSer 852
Db 2599 TCCAGGCTTGCATAATTTACCCAGATATAGCAATCTTGGGACAGGAATAATATATCC 2658
Qy 853 LeuIleTyGlyTrpGlnLeuThrLeuLeuLeuAlaIleValProIleIleAlaIle 872
Db 2659 TTAATCTATGTTGGCAACTGACACTGTTACTCTTAGCAATTTAGCCCATCATTTGCAATA 2718
Qy 873 AlaGlyValValGluMetLysMetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeu 892
Db 2719 GCAGGAGTGTTCGAATGAAATTTGTCCTGGACAGCACTGAAAGATAGAAAGAACTA 2778
Qy 893 GluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeu 912
Db 2779 GAAGGTGCTGGGAAGATCGCTACTGAAGCAATAGAAAACCTTCGGAACCTGTTGTTCTTTG 2838
Qy 913 ThrArgGluGlnLysPheGluTyMetTyAlaGlnSerLeuGlnValProTyArgAsn 932
Db 2839 ACTCAGGACGAGAAGTTTGAACATATGATGATCAGAGTTTGCAGGTACCATACAGAAAC 2898

QY 933 SerLeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyr 952
|||||
Db 2899 TCTTTGAGGAAACACACATCTTTTGGATCATCGTTTCTTCACGACGCAATGATGAT 2958
QY 953 PheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsn 972
|||||
Db 2959 TTTTCTATGCTGGATGTTCGGGTTTGAGGCTACTTGGTGGCACATGATCATGAGC 3018
QY 973 PheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGln 992
|||||
Db 3019 TTTGAGGATGTTCTGTAGTATTTTTCAGCTGTGTCTTTGGTGCATGCCGTGGGGCAA 3078
QY 993 ValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMet 1012
|||||
Db 3079 GTCAGTTTCATTTGCTCTGACTATGCGAAAGCCAAAGATATCCAGCCACATCATCATG 3138
QY 1013 IleIleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThr 1032
|||||
Db 3139 ATCATTGAAAAACCCCTTTGATTGACAGCTACAGCACAGAAGGCTAAAGCCGAACACA 3198
QY 1033 LeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIle 1052
|||||
Db 3199 TTTGAAAGGAATGTACATTTTAAAGAAGTTGTATTCAACTATCCACCCGACTGGACATC 3258
QY 1053 ProValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrIleAlaLeuValGly 1072
|||||
Db 3259 CCAGTCTTCAGGGGCTGAGCTGGAGTGAAGAGGGCCACAGCTGGCCCTGGTGGGC 3318
QY 1073 SerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeu 1092
|||||
Db 3319 AGCAGTGGCTGGGAAGACACGGTGGTCCAGCTCCTGGAGCGGTCTTATGACCCCTTG 3378
QY 1093 AlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArg 1112
|||||
Db 3379 CGGGGAAAGTGTCTGTACGGCAAGAAATAAACAACACTGAATGTTTCGTCGCTCCA 3438
QY 1113 AlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsn 1132
|||||
Db 3439 GCACACTGGGCATCGTGTCCAGAGGCCATCTCTTTGACTGCAGCATTAGTGAGAAC 3498
QY 1133 IleAlaTyrGlyAspAsnSerArgValValSerHisGluGluIleMetGlnAlaAlaLys 1152
|||||
Db 3499 ATTGCTATGGAGAACACACCGGGTGGTGCACAGGAAGAGATCGTGAGGGCAGCCAAG 3558
QY 1153 GluAlaAsnIleHisHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGly 1172
|||||
Db 3559 GAGGCCAATATACACCGCTTCATCGAGTCACTGCCTTAATAATATATACACAGAGTAGGA 3618
QY 1173 AspLysGlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeu 1192
|||||
Db 3619 GACAAAGAACTCAGCTCTCTGTGGGCCAGAAACACGCAATTCGCATAGCTCGTGCCCTT 3678
QY 1193 ValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer 1212
|||||
Db 3679 GTTAGACAGCGCTCATATTTTGGATGGAAGCCACATCATCATCTCTGGATACAGAAAGT 3738
QY 1213 GluLysValValGlnGlnAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIle 1232
|||||
Db 3739 GAAAGGTTGTCCAAAGCCCTGGCAAAAGCCAGAGAGCGCGTACCTGCATGTGTGATT 3798
QY 1233 AlaHisArgLeuSerThrIleGlnAsnAlaAspLeuIleValValPheGlnAsnGlyLys 1252
|||||
Db 3799 GCTCAGCGCTGTCCACCATCCAAATGCAGACTTATAGTGTGTGTTTCCAGAAATGAGCA 3858
QY 1253 ValLysGluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMet 1272
|||||
Db 3859 GTCAAGGAGCAGCGCACATCATCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATG 3918
QY 1273 ValSerValGlnAlaGlyAlaLysArg 1281
|||||
Db 3919 GTCAGTGTCCAGGTGGAGCAAGCGC 3945

RESULT 10

AAZ49333
ID AAZ49333 standard; cDNA; 3860 BP.
XX
AC AAZ49333;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human G185V mutant multidrug resistance-1 (MDR-1) cDNA.
XX
KW Multidrug resistance; MDR-1; P-glycoprotein;
KW transmembrane efflux pump; haematopoietic stem cell; transduction;
KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW gene therapy; gene replacement; genetic defect; thalassaemia;
KW Gauchier's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW cytokine; mutant; ds.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 1..3843
/*tag= a
/product= "Human G185V mutant MDR-1 protein"
XX
FN WO9961589-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11825.
XX
PR 28-MAY-1998; 98US-0086988.
XX
(SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX Sorrentino B, Bunting K;
XX
XX WPI; 2000-072615/06.
XX P-PSDB; AAY58187.
XX
XX Ex vivo expansion of hematopoietic stem cells transduced with a
sequence encoding human multidrug resistance-1, used for bone marrow
transplantation -
XX
XX Example 1; Page 79-82; 113pp; English.
XX
XX This sequence represents cDNA encoding human G185V mutant multidrug
resistance protein MDR-1, where the Gly residue at position 185
of the wild-type protein (AAY58186) is replaced by Val. MDR-1 is a
transmembrane efflux pump, responsible for the export of drugs from
cells, particularly cancer cells. The wild-type MDR-1 shows increased
resistance to etoposide and decreased resistance to vinca alkaloids
compared with the G185V mutant. The invention relates to transducing
haematopoietic stem cells with nucleic acid encoding an MDR protein
and culturing the modified cells. The modified haematopoietic stem
cells are useful in bone marrow transplantation (to reconstitute
haematopoietic systems in patients who have undergone chemotherapy or
radiation therapy) and in ex vivo gene therapy of genetic defects in
cells derived from haematopoietic stem cells, e.g., thalassaemia,
Gauchier's disease, sickle cell anaemia or leukaemia. The modified
cells can also be used to identify factors involved in regulating
proliferation and differentiation in haematopoietic stem cells.
XX Haematopoietic stem cells that express MDR-1 will be protected against
chemotherapeutic agents, so can be engrafted while the patient is
undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
provides sufficient cells to permit standard biochemical analysis.
XX Overexpression of MDR-1 allows cytokine-driven expansion of
haematopoietic stem cells by at least 10-fold compared with a maximum
of 4-fold in known procedures.
XX
SQ Sequence 3860 BP; 1135 A; 746 C; 956 G; 1023 T; 0 other;

Alignment Scores: 0 Length: 3860
Pred. No.:

[illegible]

Db 2032 CAAGACAGAAACCTTAGTACCAAAAGAGGCTCTGGATGAAGATATACCTCCAGCTTTCCTTT 2091
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2092 TGGAGGATTATGAAGCTAAATTAACCTGAATGGCCCTATTTTGGTTGGTGTATTTTGT 2151
Qy 720 AlaIleIleAsnGlyClyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2152 GCCATTATAAATGGAGGCTGCAACAGCATTTGCAATAAATTTTCAAGAAATATAGG 2211
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2212 GTTTTTCACGAATTCATGATCCTGAAACAAACGACAGAAATAGTAACCTGTTTTCAC 2271
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2272 TTTCTTCTAGCCCTTGGAAATATTTCTTTTATATATTTTCTTCAAGGTTTTCATATT 2331
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2332 GGCAAAGCTGGAGAGATCCTCACCAAGCGCTCCGATACATGGTTTCCGATCCATGCTC 2391
Qy 800 ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2392 AGACAGATCTGAGTTGGTTGATGACCTTAAACACACACTGGAGCATTTGACTACGAG 2451
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2452 CTCGCCAATGATGCTCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTCTGCTGAATTAC 2511
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2512 CAGAAATACCAATCTTGGCAGCAGGAATAATATATCCTTCATCTATGTTGGCAACTA 2571
Qy 860 ThrLeuLeuLeuAlaIleValProIleAlaIleAlaIleAlaGlyValValGluMetLys 879
Db 2572 ACACGTGTACTCTTAGCAATTTGACCATCATTTGCAATAGCAGGAGTTGTGAATGAA 2631
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2632 ATGTGTCTCGACAGCACHGAAGATAGAAAGACTAGAAAGTCTGGGAAGATCGCT 2691
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2692 ACTGAACAATAGAAACTTCGGAACCGTTGTTCTTGTGACTCAGGAGCAGAAAGTTGAA 2751
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2752 CATATGTATGCTCAGAGTTTGGAGGTACCATACAGAAACTCTTTGAGGAAGCACACATC 2811
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2812 TTTGGAATACATTTTCTTCCACCAGCAATGATGATATTTTCTATGCTGGATGTTTC 2871
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 2872 CGGTTTGGAGCCTACTTGGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 2931
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerPheAlaProAsp 999
Db 2932 TTTTTCAGCTGTGTCTTGGTGCCATGGCCGAGGAGTCAAGTCAATTTGCTCTGAC 2991
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleLeuLysSerProLeu 1019
Db 2992 TATGCCAAGCCAAATATACAGAGCCACATCATCATGATCATTTGAAAAACCCCTTG 3051
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3052 ATTTGACAGCTACAGCAGGAAGGCTTATGCCGAACACATTTGGAAGGAATGTCACATTT 3111
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3112 GGTGAAGTTGTATTCACTATCCACCGACCGGACATCCAGCTGCTCAGGGACTGAGC 3171

Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3172 CTGGAGTGAAGAGGCGCCAGAGCTGGCTCTGGTGGCAGCAGTGGCTCTGGGAAGAGC 3231
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3232 ACAGTGTCTCCAGCTCTCTGGAGCGGTTCTAGACCCCTTGGCAGGAAAGTGCCTGTGAT 3291
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3292 GGCAAAAGAAATAAGCCACATGAATTCAGTGGCTCGAGCACACCTGGGCATCGTGCTCC 3351
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3352 CAGGAGCCATCTCTTGTGACTGCAGCATTTGCTGAGAACATTTGCCATATGGAGACAACAGC 3411
Qy 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3412 CGGTTGGTGTCCACAGGAAGAGATCGTAGGGCAGCAAGGAGGCCAACATACATGCTCTC 3471
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3472 ATCGAGTCACTGCCTTAATAATATAGCACTAAAGTAGGAGACAAGGAAGTCACTCTCT 3531
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3532 GGTGCCCAGAAACAACGCATTTGCCATAGCTGCTGCCCTTGTAGACAGCCCTCATATTGT 3591
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3592 CTTTGGATGAAGCCAGCCAGCTCAGCTCTGGATACAGAAAGTGTCCAAGAGCC 3651
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3652 CTGGACAAGCCAGAGAGCGCCACCTGCATTTGTGATTGCTCACCGCTGTCCACCATC 3711
Qy 1240 GlnAsnAlaAspLeuIleValIlePheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3712 CAGAATGCAGACTTAATAGTGTGTTTTCAGAAATGGCAGAGTCAAGGAGCATGGCAGCAT 3771
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3772 CAGCAGCTGCTGGCAGACAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 3831
Qy 1280 LysArg 1281
Db 3832 AAGCGC 3837
RESULT 11
ABA94366
ID ABA94366 standard; DNA; 3860 BP.
XX
AC ABA94366;
XX
XX 26-MAR-2002 (first entry)
DT
XX
DE Human BCRP DNA related seq Id No. 3.
XX
XX Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;
KW cardiant; gene therapy; ds.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..3843
XX /*tag= a
PN WO200192877-A2.
XX
PD 06-DEC-2001.
XX

PF 30-MAY-2001; 2001WO-US17459.
 XX
 PR 31-MAY-2000; 2000US-0584586.
 PR 29-MAY-2001; 2001US-0866866.
 XX
 PA (SUUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 XX
 PI Sorrentino B, Schuetz J;
 XX
 DR WPI; 2002-114368/15.
 DR P-PSDB; ABB07267.
 XX
 PT Identifying a stem cell, for treating e.g., muscular dystrophy,
 PT myocardial infarction, Parkinson's disease, or neurodegenerative
 PT disorders, comprises detecting the expression of an ATP transport
 PT protein (BCRP) by a cell
 XX
 PS Disclosure; Page 59-60; 87pp; English.
 XX
 CC The invention provides a method of identifying and/or isolating a stem
 CC cell that involves detecting the expression of an ATP transport protein
 CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
 CC comprising stem cells. The isolated stem cells may be used in the
 CC treatment of diseases such as muscular dystrophy, degenerative liver
 CC disorder, myocardial infarction, Parkinson's disease, degenerative liver
 CC disorders of the brain, and for tissue regeneration or replacement.
 CC Haematopoietic cells can be used in bone marrow transplants (e.g., for
 CC treatment of leukemia) and for ex vivo gene therapy for treating blood
 CC diseases such as sickle cell anemia and thalassemia. The stem cells can
 CC also be used as cell targets in gene therapy protocols. The present
 CC sequence represents a sequence related to the BCRP for which no relevant
 CC information has been provided in the specification.
 XX

SQ Sequence 3860 BP; 1135 A; 746 C; 956 G; 1023 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 3860
 Score: 5857.00 Matches: 1163
 Percent Similarity: 95.01% Conservative: 55
 Best Local Similarity: 90.72% Mismatches: 60
 Query Match: 90.47% Indels: 4
 DB: 24 Gaps: 4

US-09-672-725C-23 (1-1281) x ABA94366 (1-3860)

QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPhetRipLysMetGly 19
 DB 1 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAGAGAACTTTTAAACTGAAC 60
 QY 20 LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
 DB 61 AATAAAGTGGAATA---GATAAGAGAGAAAGAAACCACTGTCAGTGTATTTCATG 117
 QY 40 PheArgTyrSerAsnTriPheLysAspArgLeuTyrMetLeuValGlyThrMetAlaAla 59
 DB 118 TTTTCGCTATTCAAAATGGCTTACCAAGCTGTATATGGTGGGAACTTTGGCTGCCATC 177
 QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspPhe 79
 DB 178 ATCCATGGGGCTGGAGCTTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 237
 QY 80 AlaAsnAlaGlyLysSerArgAsnLysThrPheProValIleLeuAsnGluSerIleThr 99
 DB 238 GCAAATGACGGA---AATTTAGAGATCTGATGTCAAACATCATTAATAGAGTGATATC 294
 QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
 DB 295 AATGATACAGGGTCTTCTCATGAAT---CTGGAGGAGACATGACCATATGCTCATATAT 351
 QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
 DB 352 TACATGGAAATTTGGTGTGGGTGCTGCTGCTTACATTCAGGTTTCATTTTGGTGC 411

QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
 DB 412 CTGGCAGCTGGAAGACAAATACACAAATAGAAAACAGATTTTTCATGCTATAATGCCGA 471
 QY 160 GlnGluIleGlyTyrPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
 DB 472 CAGGAGATAGGCTGGTTGGATGTCAGCATGTTGGGAGCTTAAACCCGACTACAGAT 531
 QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
 DB 532 GATGCTCTAAGATTAAATGAATTTGGTGCACAAATTTGGAATGTTCTTTCAGTCAATG 591
 QY 200 AlaThrPheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
 DB 592 GCACATATTTTCACTGGGTTTATAGTAGAGTTTACACGCTGGTTGGAAGCTAACCCCTG 651
 QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleThrPheAlaLysIleLeuSer 239
 DB 652 ATTTTGGCCATCAGTCTGCTTGGACTGCTGCTGCTGGGCAAGATATCTATCT 711
 QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
 DB 712 TCATTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTACCTGAAGGTC 771
 QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuArgTyr 279
 DB 772 TTGGCAGCAATTAAGACTGTGATTGCTTGGAGGACAAAGAAAGAACTTGAAGGTAC 831
 QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle 299
 DB 832 AACAAAAATTTAGAGAGCTAAAGAAATTTGGATTAAGAAAGCTATTACAGCCATAT 891
 QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
 DB 892 TCTATAGGTGCTGCTTTCCTGCTGATCTATGATCTTATGCTGCTGCTGCTGCTGCTG 951
 QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
 DB 952 ACCACCTTGGTCTCTCAGGGGAATATCTTATGGCAGAGTACTCAGCTATCTTTCT 1011
 QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
 DB 1012 GTATTAAATTTGGGCTTTTAGTGTGGACAGGCATCTCCAAGCATTTGAAGCATTTGCA 1071
 QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleLeuAspAsnLysProSerIleAspSer 379
 DB 1072 GCAAGAGGAGCAGCTTATGAAATCTTCAAGATAATTTGATAAAGCCAAAGTATTGAC 1131
 QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
 DB 1132 TATTCGACAGTGGGCACACCAACCATATATTAAAGGAAATTTGGAATTCGAAATGTT 1191
 QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
 DB 1192 CACTTCAGTTCACCATCTCGAAAGAAAGTAAAGATCTTGAAGGGCTTGAAGCTGAAG 1251
 QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrVal 439
 DB 1252 CAGATGGGCACACGGTGGCCCTGGTTGGAACAGTGGCTGTGGGAAGAGACACACAG 1311
 QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
 DB 1312 CAGCTGATGACAGAGCTCTATGACCCACACAGGGGATGGTGGTGTGATGGACAGAT 1371
 QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
 DB 1372 ATTAGACCATAAATGTAAGGTTTCTACGGGAAATCATTTGGTGTGGTGTGAGTACG 1431
 QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
 DB 1432 GTATTGTTGGCCACCATGAGCTGAAACATTCGCTATGCGCTGAAATGATGACCATG 1491
 QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519

1492 GATGAGATTGAGAAAGCTGTCGAAGAGCAATGCTATGACTTATCATGAAATGCCT 1551
 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
 1552 CATAAATTTGACACCCTGCTGGAGAGAGAGGGCCAGTTCAGTGGTGGGAGAAGCAG 1611
 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleIleLeuLeuAspGluAla 559
 1612 AGGATCGCCATTGCAGTGCCTGCTGCAACCCCAAGATCCTCTGCTGGATGAGGCC 1671
 560 ThrSerAlaLeuAspThrGlnSerGluAlaValGlnValAlaLeuAspLysAlaArg 579
 1672 ACCTGAGCCTTGACACAGAAAGCGAAGCAGTGGTTCAGTGGCTCGATGAAAGCCAGA 1731
 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
 1732 AAAGGTCGAGACCACCATTTGATAGCTCATCGTTTCTACAGTTGCTAAGTGCACGTC 1791
 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
 1792 ATCGCTGGTTTCGATGATGGAGTCATTTGGAGAAAGGAATCATGATGAACCTCATGAAA 1851
 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
 1852 GAGAAAGGCATTTACTTCAAACTTGTCCAAATCGACAGCAGCAAGAAATGAAGTTGAATTA 1911
 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuLeuMetSerProLysAsp 659
 1912 GAAATCGAGCTGATGAATCAAAAGTGAATTTGATGCCCTTGGAAATGCTTCCAAATGAT 1971
 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
 1972 TCAAGATCCAGTCTAAATGAAGAAAGATCAACTCGTAGGAGTTCGCTGGATCACAGCC 2031
 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
 2032 CAAGACAGAAAGCTTAGTACCAGAGAGGCTCTGGATGAAAGTATACCTCCAGCTTCCCTTT 2091
 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
 2092 TGGAGGATTATGAAGCTAAATTTAACTTGAATGGCCTTATTTGTTGTGTGATTTGT 2151
 720 AlaIleIleAsnGlyLysLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
 2152 GCCATTATAATGAGGCGCTGCACACAGCATTTCCATAATATATTTTCAAAAGATTATAGG 2211
 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
 2212 GTTTTACAGAATTTGATGATCCTGAAACAAACACAGAGATAGTAACTTGTTTTCACTA 2271
 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
 2272 TTGTTTTCAGCCCTGGAAATTTATTTCTTTTATACATTTTCTTCAAGGTTTCACATTT 2331
 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
 2332 GGCRAAGCTGGAGAGATCCTCACCAAGCGGCTCCGATACATGTTTCCGATTCATCCTC 2391
 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
 2392 AGACAGGATGTGAGTTGTTGATGACCTTAAACACACACCTGGAGCATTTGACTACACAGG 2451
 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
 2452 CTCGCCATGATGCTGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTGCCTGTAATACC 2511
 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
 2512 CAGATATAGCAAACTTTGGGACAGAAATATATATCTCATCTATGTTGGTGGCAACTA 2571
 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879

Db 2572 AACTGTGTACTCTTAGCAATTGTACCCATCATTTGCCAATAGCAGGAGTGTGTAATGAAA 2631
 QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluAlaGlyLysIleAla 899
 Db 2632 ATGTTTCTGGCAAGCACATGAAGATAAGAAAGAACTAGAAAGTGTGGGAAAGATCGCT 2691
 QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
 Db 2692 ACTGAAGCAATAGAAAACCTCCGAACGCTGTTTCTTTGACTCAGGAGCAGAAGTTTGAA 2751
 QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
 Db 2752 CATATGTATGCTCAGAGTTTCAGGTACCATACAGAAACTCTTTGAGAAAGCACACATC 2811
 QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
 Db 2812 TTTGGAATTAATTCATTTCTTCCACCAGCAATGATGATATTTTCTATCTGCTGGATGTTT 2871
 QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
 Db 2872 CGGTTTGGAGCCTACTTGGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTTAGTA 2931
 QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
 Db 2932 TTTTCAGCTGTGTCTTTGGTGCCATGCCCTGGGCAAGTCAGTTCATTTGCTCCTGAC 2991
 QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
 Db 2992 TATGCCAAAGCCAAAATATCAGCAGCCACATCATCATGATCATTTGAAAAACCCCTTG 3051
 QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
 Db 3052 ATTGACAGCTACAGCAGCGAAGGCGCTAATGCCAAACACATTGGAAAGGAATGTCACATT 3111
 QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
 Db 3112 GGTGAAGTGTATTCAACTATCCACCGCAGGACATCCAGTCTTTCAGGAGCATGAGC 3171
 QY 1060 LeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
 Db 3172 CTGGAGGTGAAGAAGGCGCAGACGCTGCTGTGGGCGAGCGTGGCTGTGGGAAGAGC 3231
 QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuLeuAsp 1099
 Db 3232 ACAGTGGTCCAGCTCCTGGAGCGGTTCACAGACCCCTTGGCAGGAGAAAGTGTGCTGTAT 3291
 QY 1100 GlyLysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSer 1119
 Db 3292 GGCAAAAGAAATAAAGCGACTGAATGTTCAGTGGCTCCGAGCACACCTGGGCGATCGTGCC 3351
 QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
 Db 3352 CAGGAGCCCATCCTGTTTGTACTGCGACATGCTGAGAACATTTGCCATGAGACACAGC 3411
 QY 1140 ArgValValSerHisGluIleMetGlnAlaLysGluAlaAsnIleHisPhe 1159
 Db 3412 CGGGTGGTGCACAGACAGAGATCGTGGGCGCAGCAAGAGGCGCAACATACATGCTCTC 3471
 QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
 Db 3472 ATCGAGTCACTGCTAATAATATAGCACTAAAGTAGGAGACAAAGAAAGTCTCAGCTCTCT 3531
 QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
 Db 3532 GGTGGCCAGAAACACGCGATGCGCATGCTGCTGCTGCTGTTAGACAGCCTCATATTTTG 3591
 QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
 Db 3592 CTTTGGATGAAGCCAGCTCAGCTCGATGATACAGAAAGTGAAGAAAGTGTCTCCAAAGGCC 3651
 QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
 Db 3652 CTGCAACAAAGCCAGAGAAGGCGCCACCTGCTGATTTGCTGCTCACCCTGCTCCACCATC 3711

QY	1240	GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis	1259
DB	3712	CAGATGACACCTTAATAGTGGTGTTCAGATGGCAGATCAAGGAGCATGGCAGCAT	3771
QY	1260	GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla	1279
DB	3772	CAGCAGCTGCTGGCACAGAAAGCATCTATTTTCAATGTCAGTGTCCAGGCTGGAACA	3831
QY	1280	LysArg	1281
DB	3832	AAGCGC	3837
RESULT	12		
AAH57442			
ID	AAH57442	standard; cDNA; 4349 bp.	
XX			
AC	AAH57442;		
DT	10-SEP-2001	(first entry)	
DE		Human intestine cell specific cDNA sequence SEQ ID NO:282.	
KW		Human; tissue specific; diagnosis; brain; heart; skeletal muscle;	
KW		lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;	
KW		metabolic disease; developmental disease; cytostatic; immunomodulatory;	
XX		neuroprotective; gene therapy; cancer; immunopathology; neuropathology.	
OS		Homo sapiens.	
XX			
PN	WO200132927-A2.		
PD			
XX	10-MAY-2001.		
PF	02-NOV-2000;	2000WO-US30396.	
XX			
PR	04-NOV-1999;	99US-0163508.	
XX			
PA	(INCY-)	INCYTE GENOMICS INC.	
PI	Sornasse T, Seilhamer JJ, Watson GA;		
XX			
DR	WPI; 2001-291057/30.		
XX			
PT	New cell and tissue specific polynucleotides useful for diagnosis,		
PT	prognosis or monitoring of treatments for disorders where the gene is		
PT	associated with a cancer, immunopathology or neuropathology -		
PS	Claim 1; Page 207-208; 327pp; English.		
XX			
CC	AAH57161 to AAH57576 represent cell and tissue specific polynucleotide		
CC	sequences (I). (I) can have cytostatic, immunomodulatory and		
CC	neuroprotective activities, and can be used in gene therapy. (I) and		
CC	proteins (II) encoded by then are used in high throughput screening		
CC	assays to select DNA molecules, RNA molecules, peptide nucleic acids,		
CC	mimetics, peptides, proteins, agonists, antagonists, antibodies or		
CC	their fragments, immunoglobulins, inhibitors, drug compounds and		
CC	pharmaceutical agents. Expression of (I) in a sample indicates the		
CC	differentiation of embryonic stem cells into a tissue selected from		
CC	brain, heart, kidney, liver, lung, skeletal muscle or pancreatic		
CC	tissues. (I) and (II) are used to produce an expression profile that		
CC	defines a metabolic or developmental process, treatment, condition,		
CC	disease or disorder. The gene profile can be used for diagnosis,		
CC	prognosis or monitoring of treatments and for investigating a		
CC	predisposition to a disorder where the gene is associated with a		
CC	cancer, immunopathology or neuropathology.		
XX			
SQ	Sequence 4349 BP; 1304 A; 816 C; 1048 G; 1181 T; 0 other;		
Alignment Scores:			
Pred. No.:	0	Length:	4349
Score:	5857.00	Matches:	1163
Percent Similarity:	95.01%	Conservative:	55

Best Local Similarity:	90.72%	Mismatches:	60
Query Match:	90.47%	Indels:	4
DB:	22	Gaps:	4
US-09-672-725C-23 (1-1281) x AAH57442 (1-4349)			
QY	1	MetAspProGluGlyGlyArgLysGlySerAla---	GluLysAsnPheTrpLysMetGly 19
DB	126	ATGGATCTTGAAGGGAGCCGCAATGGAGGAGCAAAAGAAAGAACTTTTAAATGAAC	185
QY	20	LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet	39
DB	186	AATAAAGTGRAAAA--GATAAGAGGAAAGAAACCAACTGTCAGTGTATTTCAATG	242
QY	40	PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle	59
DB	243	TTTCGCTATTCAAAATGGCTTGACAAGTGTATATGGTGGGAACTTTGGCTGCCATC	302
QY	60	IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe	79
DB	303	ATCCATGGGGCTGGACTTCCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT	362
QY	80	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr	99
DB	363	GCAATGCGAGA---AATTTAGAAGATCTGATGTCACACATCACTAATAGAGTGATATC	419
QY	100	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr	119
DB	420	AATGATACAGGGTCTCTTCATGAAT--CTGGAGGAGACATGACCAGTATGCCTATTAT	476
QY	120	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaIleGlnValSerPheTrpCys	139
DB	477	TACAGTGAATGGTGGTGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	536
QY	140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159
DB	537	CTGCAGCTGGAAGACAAATACACAAATTAGAAACAGTTTTTTCATGCTATAATCGGA	596
QY	160	GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179
DB	597	CAGGATAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	656
QY	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle	199
DB	657	GATGCTCTAAGATTAATGAAGTTATTTGGTGACAAATTTGGAATGTTCTTTCAGTCAATG	716
QY	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219
DB	717	GCAACATTTTTTCACTGGGTTTATAGTAGGATTTACACGTGGTGGAGAGCTAACCCITGTG	776
QY	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239
DB	777	ATTTTGGCCATCATGCTCTGTTGGACTGTCAGCTGCTGCTGGCAAGACTATCT	836
QY	240	SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal	259
DB	837	TCATTTACTATAAGAACTCTTAGCGTATGCAAAAGCTGGACGAGTAGCTGAAGAGGTC	896
QY	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr	279
DB	897	TTGCGACAATTAGAACTGTGATTTGGAGGACAAAAGAAAGAACTTTGAAGGTAC	956
QY	280	AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299
DB	957	AACAAAATTTAGAAGAGCTAAAGAAATTTGGGATTAAGAAAGCTATTACAGCAATATT	1016
QY	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319
DB	1017	TCATAGGTGCTGCTTCTGCTGCTGATCATGCTATATGCTCTGGCCCTCTGGTATGGG	1076
QY	320	ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer	339
DB	1077	ACCACCTTGGTCTCTCAGGGGAATATTCTATTGGACAAGTACTACTGCTATTCTTTCT	1136


```
Db 3297 CTGGAGGTGAAGAGGGCCAGACGCTGGCTGTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3356
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuValGlySerValLeuIleAsp 1099
Db 3357 ACAGTGTCCAGCTTCCTGGAGCGGTTCTACGACCCCTTGGCAGGGAAGAGTGGCTGTGAT 3416
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3417 GGCAGAGAAATAAGCGACTGAATGTTTCAGTGGCTCGAGCACACCTGGGCGATCGTGTC 3476
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3477 CAGAGGCCATCTGTTTGTAGCTCAGCATTCCTGAGAACATTCCTATGGAGACAACAGC 3536
Qy 1140 ArgValValSerHisGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3537 CGGTGGTGTGCACAGGAGAGATCGTGGAGGCAGCAAGAGGCCAACATACATACGCTTC 3596
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3597 ATCGAGTCACTGCTAATAATATAGCACATAAGTAGGAGACAAGAACTCAGCTCTCT 3656
Qy 1180 GlyClyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3657 GGTGGCCAGAAACAACGCAATGCGCATAGCTGCTGCCCTGTTTASACAGCCTCATATTG 3716
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3717 CTTTGTGATGAGCCAGCTCAGCTCTGGATACAGAGTGAAGAGTTGTCCAAAGGCC 3776
Qy 1220 LeuAspLysAlaArgGluArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3777 CTGGACAAGCCAGAGAGCCGACCTGCATTGTGCTCACCCTGCTCCACCATC 3836
Qy 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3837 CAGAATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGAGTCAAGGAGCATGGCAGCAT 3896
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3897 CACGACTGCTGGCACAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAAACA 3956
Qy 1280 LysArg 1281
Db 3957 AAGCGC 3962

RESULT 13
AA294738
ID AA294738 standard; cdna; 4646 BP.
XX
AC AA294738;
XX
DT 01-AUG-2000 (first entry)
XX
DE Human ATP binding cassette ABCB1 (MDR1) cdna.
XX
KW ABCB1; ATP binding cassette; human; cholesterol; lipid disorder;
KW atherosclerosis; lipid disorder; dyslipidemia; psoriasis;
KW lupus erythematosus; diagnosis; gene therapy; MDR1;
KW multidrug resistance; chromosome 7q21; ss.
XX
OS Homo sapiens.
XX
PN WO200018912-A2.
XX
PD 06-APR-2000.
XX
PF 21-SEP-1999; 99WO-EP06991.
XX
PR 25-SEP-1998; 98US-0101706.
XX
PA (FARB ) BAYER AG.
```

```
XX Schmitz G, Klucken J;
PI WPI; 2000-293151/25.
XX
DR Adenosine triphosphate binding proteins useful for identifying agents
XX for treating atherosclerosis and other inflammatory disorders -
PT Claim 9; Page 110-112; 154pp; English.
XX
PS
XX
```

The present sequence is that of human ATP binding cassette subfamily B protein ABCB1 cDNA. The cDNA was identified using a differential display method in which monocytes from peripheral blood were subjected to macrophage differentiation and cholesterol loading with acetylated low density lipoproteins and subsequent deloading with high density lipoprotein (HDL3) to identify cholesterol sensitive genes. The gene maps to chromosome 7q21 and is also termed MDR1 (multidrug resistance). The invention provides cholesterol-sensitive ABC genes (see AA294734-63). These genes, and polypeptides encoded by them, can be used for diagnostic and therapeutic applications, and for biochemical or cell-based assays to screen for pharmacologically active modulator compounds useful for the treatment of lipid disorders, atherosclerosis or other inflammatory diseases such as psoriasis and lupus erythematosus.

XX Sequence 4646 BP; 1371 A; 892 C; 1129 G; 1254 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 4646
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.47% Indels: 4
DB: 21 Gaps: 4

US-09-672-725C-23 (1-1281) x AA294738 (1-4646)

```
Qy 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 425 ATGGATCTTGAAGGGGAGCGCAATGGAGGAGCAAGAAGAAGAACTTTTTTAACTGAAC 484
Qy 20 LysLysSerLysLysAsnGluLysLysLysLysProThrValSerThrPheAlaMet 39
Db 485 AATAAACTGAAAA---GATAAGAGGAAAGAAACCACTGTCAGTGTATTTCAATG 541
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 542 TTTCGCTATTCAAAATGGCTTGACAAAGTTGTATATGGTGGTGGGAACCTTTGGCTGCCATC 601
Qy 60 IleHisGlyAlaAlaLeuProIleuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 602 ATCCATGGGGCTGGACTTCCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 661
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 662 GCAATGCGAGCA---AATTTAGAGACTGTGATGTCACACATCACTAATAGAGTCATATC 718
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 719 AATGATACAGGGTCTCTTCATGAAT---CTGGAGGAGACATGACCAGGTATGCTATTAT 775
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaIleGlnValSerPheTrpCys 139
Db 776 TACAGTGGAAATGGTGTGGGGTGTGCTGTCTTACATTCAGGTTTCATTTTGTGTC 835
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 836 CTGGCAGCTGGAACACAAATACACAAATAGAAAACAGTTTTTTCATGCTATATGCCGA 895
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 896 CAGGAGATAGGCTGTTTGTATGTCACGATGTTGGGAGCTTAACACCCGACTTACAGAT 955
```

QY 180 AspValSerLysIleAsnGluCluIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 956 GATGCTCTAGATTAAAGATTATTGGTGACAAAATTTGGAATGTTCTTTCACTCAATG 1015
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 1016 GCACATTTTTCACCTGGGTTTATAGTAGATTACACGTGGTTGGAAGCTTAACCTTTGTG 1075
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 1076 ATTTTGGCCATCAGTCTGTTCTTGACTGTGACTGTGCTGGCGAAAGATACTATCT 1135
QY 240 SerPheThrAspLysGluLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 1136 TCATTTACTGATAAAGAACTCTTAGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 1195
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 1196 TTGGCAGCAATTAGAACTGTGATTGCATTTGGAGGACAAAAGAAAGAACTTTGAAAGGTAC 1255
QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 1256 AACAAAAATTTAGAAAGCTTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCCAATATT 1315
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 1316 TCTATAGGTGCTGCTTCTGCTGATCTATGATCTTATGGACAAGTACTCAGCTGATTTCTT 1375
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 1376 ACCACCTTGGCTCTCGAGGGGAATTTCTATTGGACAAGTACTCAGCTGATTTCTTTTCT 1435
QY 340 ValIleIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1436 GTATTAAATTTGGGGCTTTTAGTGTGGACAGGCATCTCAAGCATTTGAAGCATTTGCCAAT 1495
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1496 GCAAGAGGAGCAGCTTATCAATCTTCAAGATAATTTGATAATTAAGCCAAAGTATTGACAGC 1555
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1556 TATTTCGAAGAGTGGGCACAAACAGATAATATTAAAGGAAATTTGGAATTCGAAATGTT 1615
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1616 CACTTCAGTTTACCATCTCGAAAAGAAAGTTAGATCTTGAAGGGCTGAACCTGAAGGTG 1675
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1676 CAGAGTGGGCAGACGCTGGCCCTGTTGGAAACAGTGGCTGTGGGAAGAGCACAAACAGTC 1735
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1736 CAGCTGATGACAGGGCTCTATGACCCACAGAGGGGATGGTCAGTGTGATGACAGAGT 1795
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1796 ATTAGGACCATAAATGTAGGTTTCTACGGGAATCATTGGTGTGGTAGTCAGGAACCT 1855
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1856 GTATTGTTGCCACCATAGCTGTAAGAAACATTGCGTATGGCGTGAAATGTCACCATG 1915
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1916 GATGAGATTGAGAAAGCTGTCAAGGAAGCCAAATGCGCTATGATCTTATCAIGAATGCGCT 1975
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1976 CATAAATTTGCACCCCTGGTTGGAGAGAGGGGCCAGTGTGAGTGGTGGGAGAGGACAG 2035

QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 2036 AGGATCGCCATTGACGTGCCCTGGTTGCAACCCCAAGATCCTCTGCTGGATGAGGC 2095
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 2096 ACGTGACCTTGGACACAGAAAGCAGAGTGGTTCAGGTGGCTCTGGATAGGCCAGA 2155
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 2156 AAAGTCGGACCCACCATTTGTATAGCTCATCGTTGTCTACAGTTCTGTAATGCTGACGTC 2215
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 2216 ATCGCTGTTTCGATGATGAGTCAATTTGTGAGAAAGAAATCATGATGAACFCATGAA 2275
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 2276 GAGAAAGCCATTTACTTCAAACTTTGCACAAATGCAGACGACGAAATGAAGTTGAATTA 2335
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuLeuMetSerProLysAsp 659
Db 2336 GAAATCCAGCTGATGAATCCAAAAGTGAATTTGATGCCCTTGGAAATGTCTTCAAAATGAT 2395
QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2396 TCAAGATCCAGTCTTAATAAGAAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAGCC 2455
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2456 CAAGACAAAAGCTTAGTACCAGAAGGCTCTGGATGAAAGTATACCTCCAGTTTCCCTT 2515
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2516 TGGAGGATTATGAAGCTTAAATTTAACTGAATGGCTTATTTTGTGTGTGTTATTTGT 2575
QY 720 AlaIleIleAsnGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2576 GCCATTATAATGAGGCGCTGCAACACCAAGCATTTGCAATAATATTTTCAAGATTATAGG 2635
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 GTTTTACAAAGATTGATGATCTCTGAAACAAACACAGACAATAGTAACTTGTTTCACTA 2695
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPheLeuGlnGlyPheThrPhe 779
Db 2696 TTGTTTCTAGCCCTTGGAAATTTATTTTATTTTACATTTTTCCTTCAGGGTTTCCACATT 2755
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 GGCAAGCTGGAGAGATCTCCACCAAGGGCTCCGATACATGGTTTCCGATCCATGCTC 2815
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2816 AGACAGGATGTGAGTTGGTTTGTATGACCTAAAAAACACCACTGGAGCATTTGACTACCAGG 2875
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2876 CTCGCCAATGATGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTGAATTACC 2935
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2936 CAGAAATAGCAAAATCTGGACAGGAATATATATCTTCATCTATGTTGGTGCACACTA 2995
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2996 ACACCTGTTTACTCTTAGCAATTTGATCCCATCATTTGCAATAGCAGGAGTTGTTGAATGAAA 3055
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 3056 ATGTTGCTGACAGACACTGAAAGATTAAGAAAGAACTAGAGGTTGCTGGGAAGATCGCT 3115
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919

Db 3116 ACTGAAGCAATAGAAACCTCCGACCCGTTGTTCTTGTACTCAGGACGACGAAGTTTGAA 3175
QY TyrMetTyraGlnSerLeuGlnValProTyraArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGTATGCTCAGAGTTTCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATC 3235
QY PheGlyValSerPheSerIleThrGlnAlaMetMetTyraPheSerTyraGlnGlyCysPhe 959
Db 3236 TTTGGAATTACATTTCTCTCACCAGGCAATGATGATATTTTCTATGCTGGATGTTTC 3295
QY ArgPheGlyAlaTyraLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 3296 CGGTTTGAGCGCTACTTGGTGGCACAATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 3355
QY PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProasp 999
Db 3356 TTTTCAGCTGTGCTTTGGTGCCATGCCGTGGGCAAGTCAGTTTCATTTGCTCCCTGAC 3415
QY TyraLysAlaLysValSerAlaAlaHisValIleMetIleGluLysSerProLeu 1019
Db 3416 TATGCCAAGCCAAAATATCAGAGCCACATCATCATGATCATTTGAAAACCCCTTGG 3475
QY IleAspSerTyraSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTGACAGCTACAGCAGCGAAGGCTTAATGCCGAACACATTTGGAAGGAATGTCACATTT 3535
QY AsnGluValValPheAsnTyraProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGTATTCAACTATCCACCCGACCGGACATCCAGTGTCTCAGGGACTGAGC 3595
QY LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3596 CTGGAGTGAAGAGGCCACACGCTGGCTCTGTGGCAGCAGTGGCTGTGGGAAGAGC 3655
QY ThrValValGlnLeuLeuGluArgPheTyraAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3656 ACAGTGTGTCAGCTCCTGGAGCGGTTCTACGACCCCTTGGCAGGGAAGTGTCTGTTGAT 3715
QY GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 GGCAAGAAATAAAGCGCACTGAATGTTTCAGTGGCTCGGACACACCTGGGCATCGTCTCC 3775
QY GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyraGlyAspAsnSer 1139
Db 3776 CAGAGGCCATCCCTGTTGACTGCAGCATGCTGTGAGAACATGTGCTATGAGAGCAACAGC 3835
QY ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3836 CGGTTGTGTACAGGAAGAGATCGTGAGGCGCAGCAAGGAGGCGCAACATACATGCTCTC 3895
QY IleGluThrLeuProGluLysTyraAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCAGTCACTGCTCCTAATAATATAGCACTAAAGTAGGACAGCAAGGAAGTCACTGCTCT 3955
QY GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GTGGCCAGAAACCAACGATTTGCCATAGCTGCTGCCCTTGTATAGACAGCCTCATATTTTG 4015
QY LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTTTGGATGAAGCCACGCTCAGCTCTGGATACAGAAAGTGAAGAGGTGTCCAGGAAGCC 4075
QY LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGGACAAAGCCAGAGAGGCGGACCTGCTGATTTGTGATTTGCTACCCGCTGTCCACCATC 4135
QY GlnAsnAlaAspIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGATGCAGACTTAATAGTGGTTTCAGAAATGGCAGAGTCAAGGAGCATGGCCGCAT 4195
QY GlnGlnLeuLeuAlaGlnLysGlyIleTyraPheSerMetValSerValGlnAlaGlyAla 1279

Db 4196 CAGCAGCTGCTGGCAGAGAAAGCATCTATTTTTCAATGTCAGTGTCCAGGCTGGAACA 4255
QY 1280 LysArg 1281
Db 4256 AAGCGC 4261
RESULT 14
AAT13394
ID AAT13394 standard; DNA; 6505 BP.
XX AAT13394;
XX 24-JUN-1996 (first entry)
XX Hybrid vector pSF-MDR.
XX Hybrid; vector; gene transfer; gene therapy; haematopoietic stem cell;
KW retroviral; murine embryonic stem cell virus; MESV;
KW Moloney murine sarcoma virus; (MOMUSV);
KW Friend murine leukaemia virus; F-MULV; ds.
XX
XX Synthetic.
XX DE19503952-A1.
XX PN
XX PD 14-MAR-1996.
XX PF 07-FEB-1995; 95DE-1003952.
XX PR 08-SEP-1994; 94DE-1431973.
XX (BOEF) BOEHRINGER MANNHEIM GMBH.
XX Baum C, Ostertag W, Stocking-harbers C, Stockingharbers C;
XX WPI; 1996-152306/16.
XX Hybrid retroviral vectors -- for gene transfer into haematopoietic
PT stem cells
XX
PS Disclosure; Page 25-29; 42pp; German.
XX New hybrid vectors comprise (1) a leader region including the U5
CC region and trna primer binding site of murine embryonic stem cell
CC virus (MESV) or Moloney murine sarcoma virus (MOMUSV), and (2) a 3'-
LTR including the U3 and R regions of a Friend murine leukaemia
CC virus (F-MULV). The vectors are useful for ex-vivo or in-vivo gene
CC therapy. High levels of gene transfer can be achieved in
CC haematopoietic stem cells and their myeloid (non-lymphatic) progeny.
CC pSF1, pSF2, pSF3 and pMM1 (sequences given in AAT13390-T13393) are
CC examples of such vectors.
CC Vector pSF-MDR (sequence given in AAT13394) is based on the
CC MESV vector R224.
XX
SQ Sequence 6505 BP; 1719 A; 1506 C; 1606 G; 1674 T; 0 other;
Alignment Scores:
Pred. No.: 0 Length: 6505
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.47% Indels: 4
DB: 17 Gaps: 4
US-09-672-725C-23 (1-1281) x AAT13394 (1-6505)
QY 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1817 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAGAACTTTTTTAACTGAAC 1876
QY 20 LysLysSerLysLysAsnGluLysLysGluLysProThrValSerThrPheAlaMet 39
Db 1877 AATAAAAGTGAAAA--GATTAAGAGGAAAGAACCAACTGTCAGTGTATTTTCAATG 1933

Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLalile 59
|||||
Db 1934 TTTGCGTATTCAAATGGCTTGACAAAGTTGTATATGGTGGGAACATTGGCTGCCATC 1993

Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
|||||
Db 1994 ATCCATGGGCGTGGACTTCTCTCATGATGCTGGTGGTGGAGAAATGACACATATCTTT 2053

Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
|||||
Db 2054 GCAATATGCCAGGA--AAITTAGAAGATCTGATGTCAAACATCATCTAAATAGAGTGATATC 2110

Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
|||||
Db 2111 AATGATACAGGGTCTTCTCATGAAT--CTGGAGGAAGACATGACCAGGTATGCTTATTAT 2167

Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 2168 TACAGTGGAAATGGTCTGGGTGCTGGTGTGCTTACATTACAGTTTCATTTTGGTGC 2227

Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 2228 CTGGCAGCTGGAGACAAATACACAAATTAGAAAACAGTTTTTCATGCTATATATGCCA 2287

Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 2288 CAGGAGATAGGCTGTTTCATGTGCACGATGTTGGGGAGCTTAACACCCGACTTACAGAT 2347

Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 2348 GATGTCTCTTAAGATTAAATGAAGTTATTGGTGACAAATTTGGAATTTCTTTTCAGTCAATG 2407

Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 2408 GCACAAATTTTCACTGGGTTTATAGTAGGATTTACAGTGGTGGAAAGCTTAACCCCTTGTG 2467

Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 2468 ATTTTGGCCATCAGTCCCTGTTCTTGAGCTGTCAGCTGCTGGCGCAAGATACATATCT 2527

Qy 240 SerPheThrAspLysGlnLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
Db 2528 TCATTTTACTTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTCT 2587

Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 2588 TTGGCAGCAATTAGAACTGTGATTGCAATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 2647

Qy 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 2648 AACAAAAATTTAGAAGAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCCAATATT 2707

Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 2708 TCTATAGGTGCTGCTTCTGCTGATCTATGATATCTTATGCTCTGGCCCTCTGGTATGGG 2767

Qy 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 2768 ACCACCTTGGTCTCTACAGGGAATATCTATTTGGACAAGTACTCCTGTTATCTTTTCT 2827

Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 2828 GTATTAAATGGGCGCTTTAGTCTGGACAGGCATCTCCAAGCATTTGAAGCATTTGCAAAAT 2887

Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 2888 GCAAGAGGAGCAGCTTATGAATCTTCAAGATAATTTGATTAATGAAGCCCAAGTATTGACAGC 2947

Qy 380 TyrSerLysSerGlyHisLysProAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 2948 TATTGCAAGAGTGGGCACAAACCAAGATATATTAGGGNAATTTGGAATTCAGAAATGTT 3007

Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 3008 CACTTTCAGTTACCCTATCTCGAAAGAAAGTTAAGATCTTTGAAGGCGCTGAAACCTCAAGGTT 3067

Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 3068 CAGAGTGGGAGAGCGGTGGCCCTGGTTTGGAAACAGTGGCTGTGGGAAGACACACACATC 3127

Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 3128 CAGCTGATGACAGAGCTCTATGACCCACAGAGGGGATGCTCAGTGTGTGACAGGAT 3187

Qy 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 3188 ATTAGGACCAATAATTAAGGTTTCTACGGGAAATCATTTGGTGTGTGAGTCAAGAACCT 3247

Qy 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 3248 GTATTGTTTGGCCACCATAGCTGAAACATTCGCTATGGCTGTGGCGTGAATAATGTCAACATG 3307

Qy 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 3308 GATGAGATTGAGAAAGCTGTCAAGGAAGCCAATGSCCTTATGACTTTATCATGAAACTGCCT 3367

Qy 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 3368 CATAAATTTGACACCCCTGGTTGGAGAGAGGGGCCAGTTGATGGTGGGCGAAGACAG 3427

Qy 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 3428 AGGATGCCATTCACCTGCGCTGGTTCGCAACCCCAAGATCCTCTGCTGGATGAGGCC 3487

Qy 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 3488 ACGTGACCCITGGACACAGAAAGCAAGCAGTGCTTCAGCTGCTGGATGAAGCCAGA 3547

Qy 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 3548 AAAGTGGGACCCACCATTTGATAGCTCATCGTTTGTCTACAGTTCGTAATGCTGACGTC 3607

Qy 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 3608 ATCCTGTTTCGATGATGAGTCAATTTGTGGAAAGGAATCATGATGAACCTCATGAAG 3667

Qy 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 3668 GAGAAAGGCATTTACTTCAAACCTTGTCACAATGTCAGACAGCAGGAAATGAAGTTGAATTA 3727

Qy 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 3728 GAAAAATGCAGCTGATGAATCCAAAAGTGAATGATGCCCTTGGAAATGCTCTCAAAATGAT 3787

Qy 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 3788 TCAAGATCCAGTCTTAATAGAAAAGATCACTCGTAGGAGTGTCCGTGGATCACAGCC 3847

Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 3848 CAAGACAGAAAGCTTAGTACCAAAAGAGGCTCTGGATGAAAGATATACCTCCAGTTTCCTTT 3907

Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 3908 TGGAGGATTTAGAGCTTAAATTTAACTGAATGGCCCTTATTTTGTGTGTGTTATTTGT 3967

Qy 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 3968 GCCATTTAATGAGGCGCTGCACACAGCATTTGCATATATATTTTCAAAGATATAGGG 4027

Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 4028 GTTTTTACAGAATTTGATGATCTCTGAAACAAACAGACAGAAATAGTAACCTGTTTCTACTA 4087

Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPheLeuGlnGlyPheThrPhe 779


```
Db 4088 TTTGTTCTAGCCCTTGGAAATATTCTTTATTACATTTTCCCTCAGGGTTTCACATTT 4147
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 4148 GGCAGAGCTGGAGAGATCTCTCACCAGCGGCTCGATACATGTTTCCGATCCATCCTC 4207
QY 800 ArgGlnAspValSerTyrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 4208 AGACAGGATGTGAGTGTGGTTGATGACCCCTTAAACACCACTGGAGCATTTGACTACCAGG 4267
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 4268 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCTGTAATTACC 4327
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTyrPgnLeu 859
Db 4328 CAGAAATAGCAATCTTGGACAGGAATAATTATATCTTCATCTATGTTGGCAACTA 4387
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 4388 ACACGTGTACTCTTAGCAATTTGACCCATCATTCGATACGAGGATTTGTTGAATGAAA 4447
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGlyLeuGluGlyAlaGlyLysIleAla 899
Db 4448 ATGTTGCTCGACAGCACTGAAGATAAGAAAGAACTAGAAAGTGTCTGGGAAGATCGCT 4507
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 4508 ACTGAACATAGAAAACCTCCCGAACCGTGTTCCTTTGACTCAGGAGCAGAAAGTTTGA 4567
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 4568 CATATGTATGCTCAGAGTTTGCAGGTACCATACAGAACTCTTTGAGGAAGCACACATC 4627
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 4628 TTTGGAAATTAATTTTCCCTTCAACCGCAATGATGATTTTCCCTATGCTGATGTTTC 4687
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 4688 CGGTTGGAGCCTACTTGTGGGCACATAAATCATGAGCTTGTGAGGATGTTCTGTAGTA 4747
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 4748 TTTTCAGCTGTTGCTTTGTGGTCATGCGCGTGGGCAAGTCAGTTCATTGCTCCTGAC 4807
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 4808 TATGCCAAAGCCAAATATCAGCAGCCCATCATCATCATGATCATGATGATGATAAACCCTTTG 4867
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 4868 ATTGACAGCTACAGCAGGAGCCCTAATGCCGACACACATTTGGAAGGAATGTCATTT 4927
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 4928 GTGGAAGTTGTATTCAACTATCCACCCAGCGGACATCCAGTGTCTCAGGGACTGAGC 4987
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 4988 CTGGAGGTGAAGAGGCCAGACGCTGGCTGGTGGCAGCAGTGGCTGTGGGAAGAGC 5047
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 5048 ACAGTGGTCCAGCTCTCGAGCGGTCTACGACCCCTTGGCAGGGAAGTGTCTGTTAT 5107
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 5108 GGCAGAGAAATAAAGCACTGAATGTTCAAGTGGCTCCGAGCACACCTGGCATCTGTGCC 5167
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
```

```
Db 5168 CAGAGGCCCTCTCTGTTGACTCCAGCATTTGCTGAGAACATTTGCCCTATGGAGACACAGC 5227
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
Db 5228 CGGGTGGTGTACAGAGAGAGATCGTGAGGCGCAAGAGAGGCCAACATACATGCTCTC 5287
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 5288 ATCGAGTCTACTGCTTAATAATATATACACTAAAGTAGGACAGAAAGAACTCAGCTCTCT 5347
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 5348 GGTGGCCAGAAACAACGATTTGCCATAGCTCTGCGCTTGTAGACAGCCTCATATTTC 5407
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 5408 CTTTGGATCAAGCCACGTCAGCTCGATACAGAAAGTGAAGGTTGTCACCAAGAGCC 5467
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 5468 CTGGACAAAGCCAGAGAGCGCCACCTGCAATTTGATTTGCTCAGCGCTGTGCCACATC 5527
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 5528 CAGAATGCAGACTTAATAGTGTGTTTCAGAAATGGCAGAGTCAAGAGCATGGCAGCAT 5587
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 5588 CAGCAGCTCTGGCAGAGAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 5647
QY 1280 LysArg 1281
Db 5648 AAGCGC 5653
RESULT 15
AAZ24041
ID AAZ24041 standard; cdna; 8630 BP.
XX
AC AAZ24041;
XX
DT 04-FEB-2000 (first entry)
XX
DE Retroviral M4 mdr-1 cDNA.
XX
KW Retroviral vector; gag gene; gene therapy; chemotherapeutic agent;
KW hematopoietic stem cell transformation; mdr-1; ss.
XX
OS Retrovirus.
XX
PN EP955374-A2.
XX
PD 10-NOV-1999.
XX
PF 07-MAY-1999; 99EP-0250151.
XX
PR 08-MAY-1998; 98DE-1022115.
XX
PA (PETT-) PETTE INST HEINRICH.
XX
PI Osterlag W, Baum C, Hildinger M;
XX
DR WPI; 2000-001087/01.
XX
PS New retroviral vector containing minimal or no gag gene sequence, for
use e.g. in gene therapy or cloning
XX
PS Disclosure; Page 16-18; 35pp; German.
XX
CC This invention describes a novel retroviral vector (RV) containing a gag
CC gene fragment having fewer than 400 bp. RV are used: (i) in gene therapy;
CC (ii) for cloning genes; (iii) for (over) expression of proteins or RNAs
CC and (iv) for transfection of hematopoietic stem cells (especially to
CC impart resistance to chemotherapeutic agents). Because of their reduced
```

CC content of viral genes, RV are very safe (no expression of toxic or
CC immunogenic proteins, no recombination with other viruses), have
CC increased cloning capacity and express non-viral sequences at a high
CC level. This sequence encodes a retroviral M4 mdr-1 protein which is
CC described in the method of the invention.

XX
SQ Sequence 8630 BP; 2235 A; 2011 C; 2187 G; 2197 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 8630
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.47% Indels: 4
DB: 21 Gaps: 4

US-09-672-725C-23 (1-1281) x AA224041 (1-8630)

QY 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
DB 1220 ATGGATCTTGAAGGGACCGCAATGAGGAGCAAGAAGAACTTTTAAACTGAAC 1279
QY 20 LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
DB 1280 AATAAAGTGAAAA--GATAAGAGGAAAGAAACCAACTGTCAGTGTATTTCAATG 1336
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLalile 59
DB 1337 TTTTCGCTATTCAAATGGCTTGCAAGTTGTATATGGTGGGAACTTTGGCTGCCATC 1396
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
DB 1397 ATCCATGGGCTCGACTCTCTCATGATGCTGTGTGGAGAAATGACAGATATCTTT 1456
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
DB 1457 GCAAAATGCAGGA---AATTTAGAAGATCTGATGTCAAACATCACTAATAAGAAGTATATC 1513
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrTrpValAlaTyrTyr 119
DB 1514 AATGATACAGGGTCTTCATGAAT--CTGGAGGAGACATGACAGATATGCTATTAT 1570
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
DB 1571 TACAGTGGAAATGGTGTGGGGTCTGTGCTGCTTACATTCAGGTTCATTTTGGTGC 1630
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
DB 1631 CTGGCAGCTGGAAGACAAATACACAAATTAGAAAACAGTTTTTTCATGCTATAATGCCA 1690
QY 160 GlnGluIleGlyTyrPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
DB 1691 CAGAGATAGCTGTTGTATGTCAGCATGTTGGGAGCTTAACCCGACTTACAGAT 1750
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
DB 1751 GATCTCTTAAGATTAAATGAATATTGGTGCACAAATTTGGAATTTCTTTCAGTCAATG 1810
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
DB 1811 GCAACATTTTCTACTGGGTTATAGTAGGATTTACAGTGGTGGAGCTTAACCCCTTGTG 1870
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
DB 1871 ATTTTGGCCATCAGCTCTGTTTGGACTGTCAGCTGCTGCTGGGCAAGACTATCT 1930
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
DB 1931 TCATTTACTGATAAAGAACTTTAGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 1990
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
DB 1991 TTGCCAGCAATTAGNACTGTGATTTGCATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 2050

QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle 299
DB 2051 AACAAAAATTTAGAGAAGCTAAAAGAAATTTGGGATAAAGAAAGCTATTACAGCAATATT 2110
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrGly 319
DB 2111 TCTATAGGTGCTGCTTTCTCTGCTGATCTATGCACTTATGCTCTGGCCTTCTGGTATGGG 2170
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
DB 2171 ACCACCTTGGTCTCTCAGGGGAATATCTATTGGACAAGTACTACTGTTCTTTCT 2230
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
DB 2231 GTATTAATTGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAT 2290
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
DB 2291 GCAAGAGGAGCAGCTTATGAAATCTTCAAGATAATTTGATATAAGCCCAAGTATTGACAGC 2350
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
DB 2351 TATTCGAAGAGTGGCACAAACAGATAATAATTAAAGGAAATTTGGAATTCAGAAATGTT 2410
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
DB 2411 CACTTCAGTTACCCTATCTCGAAAAAGAGTTAAGATCTTGAAGGGCCTGAACCTGAAGGTG 2470
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
DB 2471 CAGAGTGGGACAGCGGTGGCCCTGTTGGAAACAGTGGCTGTGGGAAGAGCACAAAGTC 2530
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
DB 2531 CAGCTGATGACAGGCTCTATGACCCACACAGGGGATGGTGCAGTGTGATGGACAGGAT 2590
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
DB 2591 ATTAGGACATAAATGTAAGGTTTCTACGGGAAATCATTTGGTGTGGTAGTCAGGAACCT 2650
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
DB 2651 GTATTTGTTGCCACCCAGTACCTGAAACATTCGCTATGCGCGTGAAATGTCCACATG 2710
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
DB 2711 GATGAGATTGAGAAAGCTGTCAAGGAAGCCAATGCTATGACTTTATCATGAACTGCCT 2770
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
DB 2771 CATAAATTTGACACCCTGTTGGAGAGAGAGGGGCCCAAGTTCCTCTCTGCTGGATGAGGCC 2830
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
DB 2831 AGGATGCCCATTTGCAGTGCCTGTTGCGAACCCCAAGATCCTCTCTGCTGGATGAGGCC 2890
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
DB 2891 ACCTCAGCCTTGGACACAGAAAGCGAGTGGTTCAGGTGGCTCTGGATAAGGCCAGA 2950
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
DB 2951 AAAGGTGCGACCAACCATTTGTGATAGCTCATCTGTTGTCTACAGTTCTGTAATGCTGACGTC 3010
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
DB 3011 ATCGCTGGTTTCGATGATGAGTCATTTGTGGAGAAAGGAAATCATGATGAATCATGAAA 3070
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
DB 3071 GAGAAAGGCATTACTTCAAACTTGTGCACATGACAGACAGAGAAATGAAGTTGAAATTA 3130

QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
|||||
Db 3131 GAAAATGCGAGCTGATGATCCAAAAGTGAAATGATGCTTGGAAATGCTCTCAATGAT 3190
QY 660 SerGlySerSerLeuIleLysArgSerThrArgSerIleHisAlaProGlnGly 679
|||||
Db 3191 TCAGATCCAGCTCTAATAAGAAAGATCAACTCGTAGGAGTGCCGCTGGATCACAAGCC 3250
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
|||||
Db 3251 CAAGACAGAAGCTTAGTACCAGAAGAGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 3310
QY 700 TrpArgIleLeuLysLeuAsnSerThrGlnTrpProTyrPheValValGlyIlePheCys 719
|||||
Db 3311 TGGAGGATATGAAGCTAAATTAACCTGAATGGCTTATTTTGTGTTGGTGTATTTGT 3370
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleGly 739
|||||
Db 3371 GCCATTATAAATGGAGGCTCTCAACAGCATTTGCAATAATATTTTCAAGATTTATAGG 3430
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
|||||
Db 3431 GTTTTACAGAATGATGATCCTGAAACAAAACGACAGATAGTAACTGTTTCTACTA 3490
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPheLeuGlnGlyPheThrPhe 779
|||||
Db 3491 TTGTTCTAGCCCTTGGAAATATTTCTTTTATTACATTTTCCTTCAAGGTTTCCACATT 3550
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
|||||
Db 3551 GGCAAAAGTGAGAGATCTCTCACCAGCGGCTCCGATACATGTTTCCGATCCATGCTC 3610
QY 800 ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
|||||
Db 3611 AGACAGGATGTAGTGGTTTGATGACCTTAAACACCACCTGGAGCATGACTACCAGG 3670
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
|||||
Db 3671 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGCTGTAATTACC 3730
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
|||||
Db 3731 CAGAATATAGCAAAATCTGGGACAGGAATTAATATATCTTCATCTATGTTGGCAACTA 3790
QY 860 ThrLeuLeuLeuLeuAlaIleValProIleAlaIleAlaIleAlaGlyValValGluMetLys 879
|||||
Db 3791 ACATGTTACTCTTAGCAATGTACCATTCATTCATAGCAGGAGTGTGTAATGAAA 3850
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluAlaGlyLysIleAla 899
|||||
Db 3851 ATGTTGCTGGACAAGCACATGAAAGATAAGAAAGAACTAGAGGTGCTGGGAAGATCGCT 3910
QY 900 ThrGluAlaIleGluAsnPheArgThrValIleSerLeuThrArgGluGlnLysPheGlu 919
|||||
Db 3911 ACTGAAGCAATAGAAAATCTCCGAACCGTGTGTTTCTTGTGACTCAGGAGCAGAAAGTTGAA 3970
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
|||||
Db 3971 CATATGATGCTCAGATTTGACGTTACCATACAGAAACTCTTTGAGGAAAGCACATC 4030
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 4031 TTTGGAATTTACATTTTCTTCCACCAGCAATGATGATTTTCTCTATGCTGGATGTTTC 4090
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
|||||
Db 4091 CGGTTTGGAGGCTACTTGGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 4150
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlnValSerSerPheAlaProAsp 999
|||||
Db 4151 TTTTCAGCTGTGTGCTTGGTGCCATAGCCGTGGGGCAAGTCAGTTTCATTTGCTCCGAC 4210
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019

Db 4211 TATGCCAAAGCCAAAATATCAGCAGCCACATCATCATGATCATTTCAAAAAACCCCTTG 4270
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
|||||
Db 4271 ATTGACAGCTACAGCAGCGAAGCCCTAATGCCGAACACATTTGGAAGAAATGTCACATTT 4330
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
|||||
Db 4331 GGTGAAGTTGTTCAACTATCCACCCGACCGGACATCCAGTGTTCAGGGACTGAGC 4390
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValIglySerSerGlyCysGlyLysSer 1079
|||||
Db 4391 CTGGAGGTGAAGAAGGCCAGACGCTGGCTCTGGTGGCAGCAGTGGCTGTGGGAAGAGC 4450
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProIleAlaGlySerValLeuIleAsp 1099
|||||
Db 4451 ACAGTGGTCCAGTCTCTGGAGCGGTTCTACGACCCCTTGGCAGGAAAGTGTGCTTGTAT 4510
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
|||||
Db 4511 GGCAAGAATAAAGCGACTGAATGTTCACTGGCTCCGAGCACACCTGGCATCGTGTCC 4570
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
|||||
Db 4571 CAGGAGCCCATCCTGTTGACTGCAGCATTTGCTGAGAACAATTCCTATGGAGACAACAGC 4630
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
|||||
Db 4631 CGGGTGGTGTACAGGAAGAGATCGTGAGGGCAGCAAGAGGAGGCAACATACATGCTTC 4690
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValIglyAspLysGlyThrGlnLeuSer 1179
|||||
Db 4691 ATCGAGTCACTGCTTAATAATATACACTAAAGTAGGAGACAAAGAACTCAGCTCTCT 4750
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
|||||
Db 4751 GGTGGCCAGAAACAAGCGCATTTGCATAGCTCGTCCCTTGTAGACAGCCTCATATTTTG 4810
QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
|||||
Db 4811 CTTTTGGATGAAGCCCGTCAGCTCTGGATACAGAAAGTGTGTCAGGAAGGCC 4870
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 4871 CTGGACAAAGCCAGAGAGGCCCGCCTGCATTTGATTTGCTCACCGCCTGTCCACCATC 4930
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||
Db 4931 CAGAATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGAGTCAAGGAGCATGGCAGCAT 4990
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 4991 CAGCAGCTGCTGGCAGAGAAAGCATCTATTTTCAATGGTGTGCTCCAGGCTGGAACA 5050
QY 1280 LysArg 1281
|||||
Db 5051 AAGCGC 5056

Search completed: November 6, 2002, 19:16:47
Job time : 589.645 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: November 6, 2002, 18:39:50 ; Search time 58.556 Seconds
(without alignments)
5373.600 Million cell updates/sec

Title: US-09-672-725C-23
Perfect score: 6474
Sequence: 1 MDPGGRKGAENFKMGK.....LLAQKGYFMSVQAGAKR 1281

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=framed_p2n_model -DEV=xlh
-Q/cgn2_1/USFTO_spool/US09672725/runat_04112002_124339_17887/app.query.fasta_1.6435
-DB-issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725.ecgn.1.134_runat_04112002_124339_17887 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5857	90.5	4646	1 US-08-181-471-2	Sequence 2, Appli
2	5857	90.5	6505	2 US-08-793-610-5	Sequence 5, Appli
3	5857	90.5	9318	2 US-08-793-610-6	Sequence 6, Appli
4	5854	90.4	4669	6 5206352-3	Patent No. 5206352
5	5830	90.1	4669	2 US-08-752-447-1	Sequence 1, Appli
6	5815.5	89.8	4264	2 US-08-784-649A-1	Sequence 1, Appli
7	5815.5	89.8	4264	2 US-08-784-649A-5	Sequence 5, Appli
8	5797	89.5	4669	2 US-08-583-276-18	Sequence 18, Appl
9	5291.5	81.7	4233	3 US-09-120-513-1	Sequence 1, Appli
10	5291.5	81.7	4233	4 US-09-450-105-1	Sequence 1, Appli
11	3628	56.0	2726	1 US-08-461-823-1	Sequence 1, Appli
12	2512	38.8	4047	2 US-08-612-734B-1	Sequence 1, Appli

13	2482	38.3	4002	2 US-08-996-545-1	Sequence 1, Appli
14	2482	38.3	4002	2 US-08-996-545-3	Sequence 3, Appli
15	2482	38.3	4002	4 US-09-328-320-1	Sequence 1, Appli
16	2482	38.3	4002	4 US-09-328-320-3	Sequence 3, Appli
17	2446	37.8	4224	1 US-08-612-521-1	Sequence 1, Appli
18	2437.5	37.7	4800	2 US-08-612-734B-3	Sequence 3, Appli
19	2288	35.3	3924	1 US-08-395-246C-1	Sequence 1, Appli
20	2059	31.8	6143	1 US-08-612-521-3	Sequence 3, Appli
21	1876.5	29.0	3924	2 US-08-996-644-3	Sequence 3, Appli
22	1876.5	29.0	3924	3 US-09-352-552-3	Sequence 3, Appli
23	1876.5	29.0	3927	2 US-08-996-644-1	Sequence 1, Appli
24	1876.5	29.0	3927	3 US-09-352-552-1	Sequence 1, Appli
25	1855	28.7	3909	1 US-08-232-537-1	Sequence 1, Appli
26	1088	16.8	4403765	4 US-09-103-840A-2	Sequence 2, Appli
27	1088	16.8	4411529	4 US-09-103-840A-1	Sequence 1, Appli
28	778.5	12.0	2376	1 US-08-394-880B-1	Sequence 1, Appli
29	766	11.8	4781	2 US-09-001-273-1	Sequence 1, Appli
30	766	11.8	4781	4 US-08-843-459A-1	Sequence 1, Appli
31	766	11.8	4847	3 US-09-061-400-1	Sequence 1, Appli
32	754.5	11.7	5889	1 US-08-463-092B-5	Sequence 5, Appli
33	754.5	11.7	5889	2 US-08-462-109A-5	Sequence 5, Appli
34	754.5	11.7	5889	3 US-08-460-907B-5	Sequence 5, Appli
35	754.5	11.7	5889	3 US-08-463-179A-5	Sequence 5, Appli
36	754.5	11.7	5889	3 US-08-461-384B-5	Sequence 5, Appli
37	727	11.2	6140	4 US-09-439-313-536	Sequence 536, App
38	709	11.0	6082	4 US-09-439-313-535	Sequence 535, App
39	708	10.9	5232	4 US-08-972-927-1	Sequence 1, Appli
40	691	10.7	2061	1 US-08-463-092B-3	Sequence 3, Appli
41	685	10.6	5011	1 US-08-462-109A-3	Sequence 3, Appli
42	685	10.6	5011	2 US-08-460-907B-3	Sequence 3, Appli
43	685	10.6	5011	3 US-08-463-179A-3	Sequence 3, Appli
44	685	10.6	5011	3 US-08-461-384B-3	Sequence 3, Appli
45	685	10.6	5011	3 US-08-461-384B-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-181-471-2
; Sequence 2, Application US/08181471
; Patent No. 5641508
; GENERAL INFORMATION:
; APPLICANT: Li, Lingna
; TITLE OF INVENTION: METHOD FOR DELIVERING BENEFICIAL COMPOSITIONS TO HAIR FOLLICLES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas Fitting
; STREET: 12526 High Bluff Drive, Suite 300
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92130
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/181,471
; FILING DATE: 13-JAN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/041,553
; FILING DATE: 02-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: ANT0029P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-792-3680
; TELEFAX: 619-792-8477

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4646 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 425..4267

; US-08-181-471-2

Alignment Scores:

Pred. No.:	0	Length:	4646
Score:	5857.00	Matches:	1163
Percent Similarity:	95.01%	Conservative:	55
Best Local Similarity:	90.72%	Mismatches:	60
Query Match:	90.47%	Indels:	4
DB:	1	Gaps:	4

US-09-672-725C-23 (1-1281) x US-08-181-471-2 (1-4646)

QY	1	MetAspProGluGlyArgLysGlySerAla---	GluLysAsnPheTrpLysMetGly	19
DB	425	ATGGATCTTGAAGGGACCGCAATGCGAGGAGCAAGAAAGCAAGCACTTTTAAACTGAAC	484	
QY	20	LysLysSerLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet	39	
DB	485	AATAAAGTGAAATA---GATAAGAGGAAAGAAACCAACTGTCAGTGTATTTCAATG	541	
QY	40	PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle	59	
DB	542	TTTCGCTATCAAAATGGCTTGACAAAGTTGTATATGGTGGGAACTTTGGCGCCATC	601	
QY	60	IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe	79	
DB	602	ATCCATGGGGCTGGACTCTCTCATGATGCTGCTGTTGGAGAAATGACAGATATCTTT	661	
QY	80	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleLeuAsnGluSerIleThr	99	
DB	662	GCAATATGCAGGA---AATTAGAAAGATCTGATGTCAAACATCACTAATAGAGTATATC	718	
QY	100	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr	119	
DB	719	AATGATACAGGGTCTTCATGAAT---CTGGAGGAAGACATGACCCAGGTATGCTATTAT	775	
QY	120	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys	139	
DB	776	TACAGTGGAAATGGTCTGGGGTGCTGGTGTGCTGTACATTGAGTTTCATTTTGGTGC	835	
QY	140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159	
DB	836	CTGGAGCTGGAGACAAATACAAATTTAGAAAACAGTTTTTCATGCTATATAGCGA	895	
QY	160	GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179	
DB	896	CAGGAGATAGCTGGTTTGATGTGACAGATGTTGGGAGCTTAACACCCGACTTACAGAT	955	
QY	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle	199	
DB	956	GATGCTCTTAAGATTAATGAAGTATTATGGTGGACAAAATTTGGAATGTTCTTCAGTCAATG	1015	
QY	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219	
DB	1016	GCAACATTTTCTACTGGTTTATAGTAGGATTTACAGTGGTGGAGGTAAACCTTGTG	1075	
QY	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239	
DB	1076	ATTTTGGCCCATCATGCTCTTGTGGACTGTCAGCTGCTGCTGGGCAAGATACATCTCT	1135	
QY	240	SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal	259	

DB	1136	TCATTTACTATAAAGAACTCTTAGCGTATGCCAAAAGCTGGAGCAGTAGCTGAAGAGGTC	1195	
QY	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr	279	
DB	1196	TTGGCAGCAATAGAACTGTGATTCATTTGGAGGACAAAGAAAGAACTTTGAAGGTAC	1255	
QY	280	AsnLysAsnLeuGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle	299	
DB	1256	AACAAAAATTTAGAAAGAACTAAAAAGAAATTTGGGATAAAGAAAGCTATTACAGGCAATATT	1315	
QY	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319	
DB	1316	TCTATAGTGCTGCTTCTCTCTGATCTATGCATCTTATGCTGCGCTTCTGTATGGG	1375	
QY	320	ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer	339	
DB	1376	ACCACCTTGGTCTCTCAGGGGAATATCTATTGGACAAGTACTACTGATTTCTTTCT	1435	
QY	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359	
DB	1436	GTATTAATTTGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAAAT	1495	
QY	360	AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379	
DB	1496	GCAAGAGGAGCAGCTTATGAAATCTTCAAGATAATTGATAATAAGCCAGTATTGCAGC	1555	
QY	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399	
DB	1556	TATTCGAAGAGTGGCACAACACAGATAATATTAAAGGAAATTTGGAATTCAGAAATGTT	1615	
QY	400	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419	
DB	1616	CACCTCAGTTACCATCTCGAAAAAGAAATTAAGATCTTGAAGGGCCCTGAACCTGAAGGTG	1675	
QY	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal	439	
DB	1676	CAGAGTGGCAGACGGTGGCCCTGGTTGGAAACACAGTGGCTGGGAAAGAGACACACAGTC	1735	
QY	440	GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459	
DB	1736	CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGATGTCAGTGTGTGAGACAGGAT	1795	
QY	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro	479	
DB	1796	ATTAGGACCAATAATGTAAGTTTCTACGGGAAATCATTTGGTGGTGGAGTCAAGAACCT	1855	
QY	480	ValLeuPheAlaThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499	
DB	1856	GTATTTGTTGGCCACCACGATAGCTGAAACATTCGCTATGCCCTGAAAAATGTCAACATG	1915	
QY	500	AspGluIleGlyLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro	519	
DB	1916	GATGAGATTGAGAAAGCTGTCAAGGAAGCCCAATGCCATGATGACTTTATCATGAAACTGCCT	1975	
QY	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539	
DB	1976	CATAAAATTTGACACCTGGTTGGAGAGAGAGGGGCCCAAGTTGAGTGGTGGCAGAGCAG	2035	
QY	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559	
DB	2036	AGGATCGCCATGTCAGTGGCTTGGTTCGCAACCCCAAGATCCTCTGCTGATGAGGCC	2095	
QY	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579	
DB	2096	ACGTGACGCTTGGACACAGAAAGGAGCAGTGGTTTCAAGTGGCTGCTGGTAAGGCCAGA	2155	
QY	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599	
DB	2156	AAAGTTCGGACACCATTTGTGATAGTCTATGCTTGTCTACAGTTCTGTAATCTGCAGCTC	2215	
QY	600	IleAlaGlyPheAspAspGlyValIleValGluLysGlyLysHisAspGluLeuMetLys	619	

Db	2216	ATCGCTGGTTTCGATGATGGAGTGATCTTGGAGAAAGGAAATCATGATCAACTCATGAAA	2275
QY	620	GIuLySGlyILeTyrPheLySLeuValThrMetGlnThrArgGlyAsnGIuLeuLeu	639
Db	2276	GAGAAAGCATTTTACTTCAACTTGTTCACAAATGCAGACAGAGAAATGAAGTTTGAATTA	2335
QY	640	GIuAsnAlaThrGlyGIuSerLySLeuValSerGIuSerAspAlaLeuGIuMetSerProLyAsp	659
Db	2336	GAAATCGAGCTGATGAATCCAAAGTGAATGATGCTTGGAAATGCTCTCAATGAT	2395
QY	660	SerGlySerSerLeuLeuLySArgSerThrArgSerThrArgSerIleHisAlaProGlnGly	679
Db	2396	TCAAGATCCAGCTCTAATAAGAAAGATCAACTCTGAGAGTGTCCGTGGATCAACAACC	2455
QY	680	GlnAspArgLySLeuGlyThrLySLeuAspLeuAsnGIuAsnValProProValSerPhe	699
Db	2456	CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCTCTT	2515
QY	700	TrpArgIleLeuLySLeuAsnSerThrGIuThrProTyrPheValValGlyIlePheCys	719
Db	2516	TGGAGGATTAAGCTAAATTAACCTGAATGGCCTTATTTGTTGTTGGTGAATTTGT	2575
QY	720	AlaIleIleAsnGlyLyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	739
Db	2576	GCCATTATAATGGAGCCTGCAACAGCATTTGCAATAATATTTCAAAGATATATAGG	2635
QY	740	IlePheThrArgAspGIuAspProGIuThrLySArgGlnAsnSerAsnMetPheSerVal	759
Db	2636	GTTTTACAGATTCATGATCCTGAAACAAACGACAGATATGACTTGTGTTTCTACTA	2695
QY	760	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	779
Db	2696	TTGTTTCTAGCCCTTGGAAATATTCTTTTATTACATTTTCCCTCAGGTTTCACATTT	2755
QY	780	GlyLySAlaGlyIleLeuThrLySArgLeuArgTyrMetValPheArgSerMetLeu	799
Db	2756	GGCAAGCTGGAGAGATCTCTACCAAGCGCTCCGATACATGTTTTCCGATCCATGCTC	2815
QY	800	ArgGlnAspValSerTrpPheAspAspProLySAsnThrThrGlyAlaLeuThrThrArg	819
Db	2816	AGACAGGATGTGAGTTGGTTTCATGACCTTAAACACCACCTGGAGCATTTGACTACCAG	2875
QY	820	LeuAlaAsnAspAlaAlaGlnValLySAlaIleGlySerArgLeuAlaValIleThr	839
Db	2876	CTGCGCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCACGGCTTGCTGTAATTACC	2935
QY	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	859
Db	2936	CAGAAATAGCAAACTTGGGACAGAAATAATATCTTCATCTATGTTGGCAACTA	2995
QY	860	ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLyS	879
Db	2996	ACACTGTTACTCTAGCAATGTACCCATCATTCCAATACAGAGGTTGTGAAATGAAA	3055
QY	880	MetLeuSerGlyGlnAlaLeuLySAspLySLeuGluGlyAlaGlyLySLeuAla	899
Db	3056	ATGTTGCTCGGACACAGCACTGAAAGATAAGAAAGAACTAGAAAGCTGCTGGAGATCGCT	3115
QY	900	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLySLeuGlu	919
Db	3116	ACTGAAGCAATAGAAACTTCGAACCGTTGTTCTTTGACTCAGACAGAGAAGTTTGA	3175
QY	920	TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLySAlaHisIle	939
Db	3176	CATATGATGCTCAGAGTTTCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATC	3235
QY	940	PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe	959
Db	3236	TTTGGAAATTACATTTTCTCCACCAGCAATGATGATATTTTCCCTATGCTGATGTTTC	3295
QY	960	ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal	979
Db	3296	CGTTTGGAGCCACTCGGTGGCACATAACTCATGAGTTTGGAGATCTCTGTTAGTA	3355

RESULT 2
US-08-793-610-5
; Sequence 5, Application US/08793610
; Patent No. 5856744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARBERS, Carol
; APPLICANT: OSTERTAG, Wolfram


```

/ TITLE OF INVENTION:  RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
/ TITLE OF INVENTION:  FOR GENE TRANSFER
/ NUMBER OF SEQUENCES:  6
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram LLP
/ STREET:  655 Fifteenth Street N.W. Suite 330
/ CITY:  Washington
/ STATE:  D.C.
/ COUNTRY:  U.S.A.
/ ZIP:  20005-5701
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE:  Floppy disk
/ COMPUTER:  IBM PC compatible
/ OPERATING SYSTEM:  PC-DOS/MS-DOS
/ SOFTWARE:  PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER:  US/08/793,610
/ FILING DATE:  07-MAR-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:  DE P 44 31 973.8
/ FILING DATE:  08-SEP-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:  DE 195 03 952.1
/ FILING DATE:  07-FEB-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:  PCT/EP95/031175
/ FILING DATE:  10-AUG-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME:  Berman, Richard J.
/ REGISTRATION NUMBER:  39,105
/ REFERENCE/DOCKET NUMBER:  P1614-7007
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE:  (202)638-5000
/ TELEFAX:  (202)638-4810
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH:  6505 base pairs
/ TYPE:  nucleic acid
/ STRANDEDNESS:  double
/ TOPOLOGY:  circular
/ MOLECULE TYPE:  DNA
/
/ US-08-793-610-5
/
/ Alignment Scores:
/ Pred. No.:  0 Length:  6505
/ Score:  5857.00 Matches:  1163
/ Percent Similarity:  95.01% Conservative:  55
/ Best Local Similarity:  90.72% Mismatches:  60
/ Query Match:  90.47% Indels:  4
/ DB:  2 Gaps:  4
/
/ US-09-672-725C-23 (1-1281) x US-08-793-610-5 (1-6505)
/
/ Qy  1  MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPhetTpLysMetGly 19
/
/ Db  1817  ATGGATCTTGAGGGGACCGCAATGGAGGACAAAGAGAGAACTTTTAAACTGAAC 1876
/
/ Qy  20  LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
/
/ Db  1877  AATAAAAGTCAAAAA---GATAAGAGGAGAAACCAACTGTCACTGTATTTTCAATG 1933
/
/ Qy  40  PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
/
/ Db  1934  TTTCCGCTATTCAATTTGGCTTGACAAAGTTGTATATGGTGGGGGAACCTTTGGCTGCCATC 1993
/
/ Qy  60  IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
/
/ Db  1994  ATCCATCGGGCTGGACTTCCTCTCATGATGCTGCTTTGGAGAAATCACAGATATCTTT 2053
/
/ Qy  80  AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
/
/ Db  2054  GCAAAATCGAGGA---AATTATGAGAAGATCTGATGTCAAACATCATCAANTAGAAGTGATATC 2110

```

[illegible]

Db	4268	CTCGCCAATGATGCTCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTATTAATACC	4337
Qy	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	859
Db	4328	CAGAAATATAGCAATCTTGGGACAGGAATAATATATCTTCATCTATGGTTGGCAACTA	4387
Qy	860	ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetIys	879
Db	4388	ACACTGTTACTTCATTAGCAAAATGTACCACATCATGCAATAGCAGGAGTTGTTGAAATGAA	4447
Qy	880	MetLeuSerGlyGlnAlaLeuLysAspIysLysGluLeuGluGlyAlaGlyLysIleAla	899
Db	4448	ATGTTGCTGTGACCAAGCACTGAAAGATAAGAAAGAACTAGAAAGTGTGGGAAGTCGCT	4507
Qy	900	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu	919
Db	4508	ACTGAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAAAGTTTGAA	4567
Qy	920	TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle	939
Db	4568	CATATGATGTCTAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAGCACACATC	4627
Qy	940	PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe	959
Db	4628	TTTGGAATTTACATTTTCTCTTACCCAGGCAATGATGATTTTCTATGCTGGATGTTTC	4687
Qy	960	ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal	979
Db	4688	CGGTTTGGACCTACTTGGTGGCACATAACTCATGAGCTTTGAGGATGTTCTGTAGTA	4747
Qy	980	PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp	999
Db	4748	TTTTTCAGCTGTGTCTTTGGTGTGCATGGCCGTGGGGCAAGTCAGTTCATTTGCTCCTGAC	4807
Qy	1000	TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu	1019
Db	4808	TATGCCAAGCCAAATATATCAGACGCCCATCATCATGATCATTTGAAAAGACCCCTTG	4867
Qy	1020	IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe	1039
Db	4868	ATTTCAGAGCTACAGCAGCGGAAGCCTAATGCCGACACATTTGGAAGAAATGTACATTT	4927
Qy	1040	AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer	1059
Db	4928	GGTGAAGTTGTATTTCAACTATCCACCCGACCCGACATCCAGTGTCTCAGGACTGAGC	4987
Qy	1060	LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer	1079
Db	4988	CTGGAGGTGAAGAAGGGCCAGACGCTGGGTCTGGTGGGAGCAGTAGTGGCTGTGGGAAGAGC	5047
Qy	1080	ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp	1099
Db	5048	ACAGTGGTCCAGCTCTCGGAGCGGTCTACGACCCCTTGGCAGGGAAAGTGCTGCTTGAT	5107
Qy	1100	GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer	1119
Db	5108	GGCAAGAAATAAAGCACTGAATGTTCAAGTGGCTCCGAGCACACCTGGGCATCGGTGCC	5167
Qy	1120	GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer	1139
Db	5168	CAGGAGCCCATCTGTTTCACTGCACATTTGCTGAGAACATTTGCCATGTGAGACACACAGC	5227
Qy	1140	ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe	1159
Db	5228	CGGGTGTCTCAGGAAGAGATCGTAGGGGACGCAAGAGAGGCCAACATACATCGCTTTC	5287
Qy	1160	IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer	1179
Db	5288	ATCAGAGTCACTGCCATAATAATATAGCACTAAAGTAGGAGACAAAGGAACCTCAGCTCTCT	5347
Qy	1180	GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu	1199
Db	5348	GGTGCCAGAAACAACGATTCGCATAGCTCGTCCCTGTTTAGACAGCCTCATATTTTG	5407

QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 5408 CTTTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCTCAAGAACCC 5467
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgIleuSerThrIle 1239
Db 5468 CTGGACAAGCCAGAGAGCCGACCTGCTGATTTGCTCACCCTGTGCCACCATC 5527
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisClyThrHis 1259
Db 5528 CAGAATGCAGACTTAATAGTGTGTTTCAGAAATGCGAGAGTCAAGGAGCATGGCCACGCAT 5587
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 5588 CACGACGCTGTCGACAGAAAGCATCTATTTTTCATGGTCAGTGTCCAGGCTGGAACA 5647
QY 1280 LysArg 1281
Db 5648 AAGCGC 5653
RESULT 3
US-08-793-610-6
; Sequence 6, Application US/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARRIS, Carol
; TITLE OF INVENTION: OSTERTAG, Wolfram
; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
; NUMBER OF INVENTIONS: FOR GENE TRANSFER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,610
; FILING DATE: 07-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 31 973.8
; FILING DATE: 08-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 195 03 952.1
; FILING DATE: 07-FEB-1995
; PRIOR APPLICATION DATA: PCT/EP95/03175
; APPLICATION NUMBER: PCT/EP95/03175
; FILING DATE: 10-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Beriman, Richard J.
; REGISTRATION NUMBER: 39,105
; REFERENCE/DOCKET NUMBER: P1614-7007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9318 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
US-08-793-610-6

Alignment Scores:

Pred. No.: 0 Length: 9318
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.47% Indels: 4
DB: 2 Gaps: 4
US-09-672-725c-23 (1-1281) x US-08-793-610-6 (1-9318)
QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1776 ATGGATCTTGAGGGGACCCCAATGGAGGAGCAAGAAAGAAAGAACTTTTAAACTGAAC 1835
QY 20 LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 1836 AATAAAAGTGAAGAA---GATAAGAGAAAGAAACCAACTGTCAGTGTATTTTCAATG 1892
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 1893 TTTGCGTATTCAAATGGCTTGACAAGTTGTATATGTTGGTGGGAACTTTGGCTGCCATC 1952
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 1953 ATCCATGGGGCTGGACTTCCTCTCATGATGCTGCTTTGGAGAAATGACAGATATCTTT 2012
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 2013 GCAAATGCAGGA---AATTTAGAAGATCTGATGCAACATCACTAATAGAGTGATATC 2069
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 2070 AATGATACAGGGTCTTCTCATGAAT---CTGGAGGAGACATGACCAGGTATGCCTATTAT 2126
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 2127 TACAGTGAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2186
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 2187 CTGGCAGCTGGAAGACAAATACACAAATATGAAACAGTGTTCATGCTTATATGCGA 2246
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGlyLeuAsnThrArgLeuThrAsp 179
Db 2247 CAGGAGATAGCTGCTGTTGATGTGCACCATGTTGGGAGCTTAACACCCGACTTACAGAT 2306
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 2307 GATGCTCTAGATTAAATGAAGTATTGTTGTCACAAAATTTGGAATGTTCTTTTCAGTCAATG 2366
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 2367 GCACATTTTTCACCTGGGTTTATAGATTTACACGCTGGTGGAGCTAACCCCTTGTG 2426
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 2427 ATTTGGCCATCAGTCTCTTCTTGGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2486
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaIleGluVal 259
Db 2487 TCATTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 2546
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 2547 TTGGCAGCAATTAGAACTGTGATTGCAATTTGGAGGCAAAAGAAAGAACTTGAAGAGTAC 2606
QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 2607 AACAAAAATTTAGAACAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCCAATAT 2666
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 2667 TCTATAGGTGCTGCTTTCTGCTGATCTATGATCTTATGCTCTGCTGCTGCTGCTGCTGCTG 2726

Db 4887 GGTGAAGTTGTATTCAACTATCCACCACCGGACATCCCGAGTCTTCAGGGACTGAGC 4946
QY 1060 LeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 4947 CTGGAGGTGAAGAGGGCCAGACGCTGGCTCTGGTGGGCACAGTGGCTGTGGGAAGAGC 5006
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 5007 ACAGTGGTCCAGCTCCTGGAGGGTCTACGACCCCTTGGCAGGGAAGTCTGCTTGAT 5066
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnThrPheArgAlaHisLeuGlyIleValSer 1119
Db 5067 GGCAAGAAGAAATAAGCGACTGAATCTTCAGTGGCTCCGAGCACACCTGGGCATCGTGCC 5126
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 5127 CAGGAGCCCTCCTGTTGACTGACGATGCTGAGACATGCTTATGGAGACAAACAGC 5186
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaLalLysGluAlaAsnIleHisPhe 1159
Db 5187 CGGGTGGTCTACAGGAAGAGATCGTGGGCGAGCAAGGAGGCCAACATACATGCCTTC 5246
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 5247 ATCGAGTCACTCCCTAATAAATATAGCACTAAAGTAGGAGACAAAGGAACCTCCTCT 5306
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 5307 GTGGCCAGAAACACGATTGCCATAGCTGCGCCTGTGTAGACACCTCATATTTTG 5366
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 5367 CTTTGGATGAAGCCACGCTCAGCTCTGGATACAGAAAGTGTAAGGTTGTCCAAGAAGCC 5426
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 5427 CTGGACAAGCCAGAGAAGCGCGCACCTGTCATTTGTGTCTACCGCCTGTCCACCATC 5486
QY 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 5487 CAGAATGCAGACTTAATAGTGGTGTTCAGAAATGCCAGATGCAGAGCATGGCAGCAT 5546
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 5547 CAGCAGCTGCTGGCAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGACA 5606
QY 1280 LysArg 1281
Db 5607 AAGCGC 5612

RESULT 4
5206352-3
; Patent No. 5206352
; APPLICANT: Roninson, Igor B.; Pastan Ira H.; Gottesman,
; Michael M.
; TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
; SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/622,836
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 892,575
; FILING DATE: 01-AUG-1986
; APPLICATION NUMBER: 845,610
; FILING DATE: 28-MAR-1986
; SEQ ID NO: 3:
; LENGTH: 4669
5206352-3

Alignment Scores:
Pred. No.: 0 Length: 4669
Score: 5854.00 Matches: 1162

Percent Similarity: 95.01% Conservative: 56
Best Local Similarity: 90.64% Mismatches: 60
Query Match: 90.42% Indels: 4
DB: 6 Gaps: 4
US-09-672-725c-23 (1-1281) x 5206352-3 (1-4669)
QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 425 ATGGATCTTCAAGGGGACCGCAATGGAGGAGCAAGAAGAAGAACTTTTAAACTGAAC 484
QY 20 LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 485 AATAAAGTCAAAA---GATAAGAAGAAAGAAACCACTGTCAGTGTATTTTCAATG 541
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLalIle 59
Db 542 TTTCCGTATTCAAATTTGGCTTGACAAAGTTGATATGGTGGGAACTTTGGCTGCCATC 601
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 602 ATCCATGGGGCTGGACTTCCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 661
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 662 GCAATGTCAGGA---AATTTAGAAGATCTGATGTCAAAACATCACTAATAAGAGTGATATC 718
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 119
Db 719 ANTGATACAGGGTCTTCATGAAAT---CTGGAGAGACATGACCAGGTATGCTATTAT 775
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 776 TACAGTGGAAATGGTGGGGTGGTGGTGTCTGCTTACATTCAGGTTCATTTTGGTGC 835
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 836 CTGGCAGCTGGAGACAAATACACAAATTTAGAAACAGTTTTTTCATGCTATAATGGCA 895
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 896 CAGGAGATAGCTGGTGTGATGTGCAGATGTTGGGAGCTTAACACCCGACTTACAGAT 955
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 956 GATGCTCTTAAGATTAATCAAGTATTGGTGACAAAATTGGAATGTCTTTCAGTCAATG 1015
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 1016 GCAACATTTTCACTGGGTTTATAGTAGGATTTACACGTTGGTGGAGAGCTAACCCCTGTG 1075
QY 220 IleLeuAlaIleSerProValLeuLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 1076 ATTTGGCCATCAGTCTCTGTTGGACTGTCAGTGTCTGCGGCAAGAGATATCTATCT 1135
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
Db 1136 TCATTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGCT 1195
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 1196 TTGGCAGCAATTAAGACTGTGATTTGCATTTGGAGACAAAGAAAGAACTTGAAGAGTAC 1255
QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 1256 AACAAAAATTTAGAAGAAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCAATAT 1315
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 1316 TCTATAGGTGCTGCTTTCCCTGATCTATGATCTATGCTCTGCTGCTGCTGCTGCTGCTG 1375
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
|||:::|||||

Db 1376 ACCACCTGGTCCTCTCAGGGGAATAATCTATTCGACAAAGTACTCACTGTATTCCTTTCT 1435
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1436 GTATTAAATGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAAAT 1495
Qy 360 AlaArgGlyAlaLaTyRGIuilePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1496 GCAAGAGGACAGCTTATGAAATCTTCAAGATATTGATTAATGAAGCAAGTATTGACAGC 1555
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1556 TATTCGAAGAGTGGCCACAACACAGATAATATTAAAGGAAATTTGGAATTCAGAAATGTT 1615
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1616 CACTTCAGTTACCATCTCGAAGAAGTTAAGATCTTGAAGGGCCGTGAACCTGAAGGTTG 1675
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1676 CAGAGTGGGACAGCGGTGGCCCTGTTGGAACAGTGGCTGTGGGAAGGACACACAGTC 1735
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1736 CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGATGGTCAGTCTTGATGGACAGGAT 1795
Qy 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1796 ATTAGGACCATAAATGTAAGTTTTCACGGGAATCATTTGGTGTGGTGAAGAACCT 1855
Qy 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1856 GTATTGTTGCCACCACGATAGCTGAAACATTCGCTATGGCGTGAAATGTCACCAATG 1915
Qy 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1916 GATGAGATTGAGAAAGCTGTCAAGGAAGCCAATGCCATGACTTTATCATGAAAGTGCCT 1975
Qy 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlnLysGln 539
Db 1976 CATAAATTTGACACCTGGTTGGAGAGAGAGGGGCCCATGTTAGTGGTGGGAGAGCAG 2035
Qy 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 2036 AGGATCGCCATTGCAGTGCCTCGTTCGCAACCCCAAGATCCTCCTGCTGATGAGGCC 2095
Qy 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 2096 ACGTCAGCCTTGGACACAGAAAGCGAAGCAGTGGTTCAGTGGCTCTGGATAAGGCCAGA 2155
Qy 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 2156 AAAGTTCGACACCATCTGTAGCTCATCTGTTTGTCTACAGTTCGTAATGCTGACGTC 2215
Qy 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 2216 ATCGCTGGTTTCGATGATGGAGTCATCTGGAGAAAGAAATCATGATGAATCATGAA 2275
Qy 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 2276 GAGAAAGGCTTTACTTCAAACTGTCAATGTCAGATGCAGACGACGAGAAATCAAGTTGAATA 2335
Qy 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2336 GAAATGACAGCTGATGATCAAAAGTGAATTTGATGCTTGGAAATGCTTCAATGAT 2395
Qy 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2396 TCAAGATCCAGTCTAATAAGAAAGATCAACTCGTAGGATGTCCGTGGATCACAAGCC 2455
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
Db 2456 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 2515

Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2516 TGGAGGATTATGAAGCTAAATTAACATGAATGCCCTATTATTTGTTGTTGTTGTTTGT 2575
Qy 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2576 GCCATTATAAATGGAGGCGCTGCACCCAGCATTTGCAATAAATATTTTCAAGAAATTATAGGG 2635
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 GTTTTTTACAAGAATTGATGATCCTGAACAAACACACAGAAATAGTAACCTGTTTCACTA 2695
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2696 TTGTTTCTTAGCCCTTGGRAATTATTTCTTTATTACATTTTTCCTTCAGGGTTTCAATTT 2755
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 GGCAAAGCTGGAGAGATCCTCACCAGCGCTCCCATGATCATGTTTCCGATGCCATGCTC 2815
Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2816 AGACAGGATGTGAGTTGTTGATCACCCCTAAAAACACCACCTGGAGCATTGACTACCAGG 2875
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2876 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTTCCAGGCTTGCCTGAATACC 2935
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2936 CAGAAATATAGCAATCTTGGGACAGAAATATATATCTTCATCTATGTTGGCAACTA 2995
Qy 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2996 ACACCTGTTTACTCTTAGCAATTTGTACCCATCATTCATATAGCAGAGTTGTTGAATGAA 3055
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 3056 ATGTTGCTGTGACACAGCACTGAAAGATAAGAAAGAACTAGAAAGTGTGGGAAGATCGCT 3115
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 3116 ACTGAAGCAATAGAAAACCTTCGGAACCGTTGTTCTTTGACTCAGGACGAGAGTTTGA 3175
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAGACACATC 3235
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTGGAATTTACATTTTCTTCACCCAGGCAATGATGATTTTCTTCTATGCTGGATGTTTC 3295
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 3296 CGGTTTGGAGCCTACTTGGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTTAGTA 3355
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3356 TTTTCAGCTGTGTCTTGTGTGTCATGGCGTGGGCAAGTCAGTTTCATTTGCTCCTGAC 3415
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAAGCCAAATATCAGCAGCCCATCATCATGATCATTTGAAAAACCCCTTTG 3475
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTCACAGCTACACACGGAAGCCTAATGCCGACACATTTGGAAGAAATGTCACATTT 3535
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGTTATTTCAACTATCCACCCGACGACATCCAGTGTCTTCAGGACTGAGC 3595

```
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3596 CTGGAGGTGAAGAAGGGCCAGACGCTGGCTCTGGTGGGCACAGTGGCTGTGGGAAGGC 3655
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3656 ACAGTGGTCCACTCTCTGGAGGGGTTCTACGACCCCTTGGCAGGGAAGTCTGCTTGAT 3715
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 GCGAAGAAATAAAGCGACTGAATGTTGAGTGGCTCCGAGCACACCTGGGCATCGTGTCC 3775
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 CAGGAGCCCATCTGTTGACCTGCAGCATTTGCTGAGACATTTGCTATGGACACACAGC 3835
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
Db 3836 CGGGTGTGTCACAGGAGAGATCGTGAGGCGCAGCAAGAGGCCAACATACATGCCCTTC 3895
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCGAGTCACTGCTTAATAATATAGCAGTAAAGTAGGAGACAAAGGAACCTCAGCTCTCT 3955
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GTGGCCAGAAACACGCAATGTCATAGCTGCTGCTGTTAGACAGCCCTCATATTTG 4015
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAAGGTTGTCTCAAGAAGCC 4075
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGGCAAAAGCCAGAGAGGCGCACCTGCATTTGATTGCTCCAGCCCTGTCACCATC 4135
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGATGCGAGACTTAATAGTGGTTTTCAGATGGCAGAGTCAAGGAGCATGGCAGCAT 4195
QY 1260 GlnGlnLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGCTGCAGAGAAAGGCATCTATTTTCAATGGTCAGTCTCCAGGCTGGAACA 4255
QY 1280 LysArg 1281
Db 4256 AAGGCC 4261

RESULT 5
US-08-752-447-1
: Sequence 1, Application US/08752447
: Patent No. 5994088
: GENERAL INFORMATION:
: APPLICANT: Meckhner, Eugene
: APPLICANT: Roninson, Igor B
: TITLE OF INVENTION: Methods and Reagents for Preparing and
: NUMBER OF INVENTIONS: Using Immunological Agents Specific for P-glycoprotein
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.
: STREET: 300 South Wacker Drive, Seventh Floor
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/752,447
: FILING DATE: 15-NOV-1996
```

```
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: No. 5994088nan, Kevin E
: REGISTRATION NUMBER: 35,303
: REFERENCE/DOCKET NUMBER: 95,1121
: TELEPHONE: 312-913-0001
: TELEFAX: 312-913-9808
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4669 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: 5'UTR
: LOCATION: 1..424
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 425..4264
: FEATURE:
: NAME/KEY: 3'UTR
: LOCATION: 4265..4669
: US-08-752-447-1
Alignment Scores:
Pred. No.: 0 Length: 4669
Score: 5830.00 Matches: 1156
Percent Similarity: 94.70% Conservative: 58
Best Local Similarity: 90.17% Mismatches: 64
Query Match: 90.05% Indels: 4
DB: 2 Gaps: 4
US-09-672-725C-23 (1-1281) x US-08-752-447-1 (1-4669)
Qy 1 MetaspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 425 ATGGATCTTGAAGGGGACCCGCAATGGAGGAGCAAGAGAGAACTTTTAAACTGAAC 484
Qy 20 LysLysSerLysLysAsnGluLysLysLysLysProThrValSerThrPheAlaMet 39
Db 485 AATAAAGTGAANA---GATAAAGAGAAAGAAACCACTGTCAGTGTATTTTCAATG 541
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 542 TTTCGCTATTCAAAATTTGGCTTGACAAGTTGTATATGTTGGTGGGAACCTTGGCTGCCATC 601
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 602 ATCCATGGGGCTGGACTTCCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 661
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 662 GCAATGTCAGGA---AATTAGAAGATCTGATGTCAAAACATCACTAATAGAAGTGATATC 718
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 719 AATGATACAGGGTTCTTCATGAAT---CTGGAGGAGAGACATGACACAGGTATGCCCTATTAT 775
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 776 TACAGTGAATTTGGTGGTGGGTGCTGGTGTGCTGCTTACATTCAGGTTTCATTTGGTGC 835
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 836 CTGGCAGCTGGAAGACAAATACACAAAATTAGAAAAACAGTTTTTTCATGCTATTATGCGA 895
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 896 CAGGAGATAGGCTGTTTGTATGTGCACGATGTGGGGAGCTTAACACCCGACTTACAGAT 955
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
```


Db	956	GATGCTCCAGATTAAAGGAATGGTGACAAATTTGGAATGCTTTCTTCAGTCAATG	1015
Qy	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyrIpLysLeuThrLeuVal	219
Db	1016	GCAACATTTTCACTGGTATTATAGTAGATTTACACGTGGTGGAGACTAACCCCTGTGTG	1075
Qy	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIlelTrpAlaLysIleLeuSer	239
Db	1076	ATTTTGGCCATCAGTCGTGTTCTTGACATGTCAGCTGCTGTGGGCAAGATACTACTCT	1135
Qy	240	SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal	259
Db	1136	TCATTTACTGATAAGAACTCTTAGCGTATGCANAAGCTGGAGCAGTAGCTGAAGAGCTC	1195
Qy	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr	279
Db	1196	TTGGCACCAATTAGAAGCTGTGATTCGATTTGGAGGACAAAGAAGAAGACTTGAAGGTAC	1255
Qy	280	AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299
Db	1256	AACAAAATTTAGAAAGCTAAAGAAATGGGATAAGAAAGCTATTACAGCCAATATT	1315
Qy	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319
Db	1316	TCATAGTGCTGCTTCTCTGCTGATCATATGCATCTTATGCTGTGGCTTCTGTGATGGG	1375
Qy	320	ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer	339
Db	1376	ACCACCTTGGTCTCTCAGGGGAATATTCATTTGGCAAGTACTCATCTGATTCTTTTCT	1435
Qy	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359
Db	1436	GTATTAAATGGGGCTTTAGTCTTGGACAGGCATCTCAAGCATTTGAAGCATTTGCAAT	1495
Qy	360	AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379
Db	1496	GCAAGAGGAGCAGCTTATGAATCTTCAAGATAATGTGATAATAGCCCAAGTATTGACAGC	1555
Qy	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399
Db	1556	TATTCGAAGAGTGGGCACAAACCAATATATTAAGGGAATTTGGAAATTTCAAGAAATGT	1615
Qy	400	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419
Db	1616	CACCTTCAGTACCATCTCGAAAGAAAGTTAAGATCTTGAAGGSCCTGAACCTGAAGGTG	1675
Qy	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrVal	439
Db	1676	CAGAGTGGGCACAGCGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGACACACAGTC	1735
Qy	440	GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459
Db	1736	CACGTGATCCAGAGCTCTATGACCCCAACAGAGGGATGGTCTAGTGTGTGATGACAGGAT	1795
Qy	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro	479
Db	1796	ATTAGGACCATAAATGTAAGGTTTCTACGGGAAATCATTTGGTGGTGGTGGAGAACCT	1855
Qy	480	ValLeuPheAlaThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499
Db	1856	GTATTGTTTGGCCACCACCATAGCTGAAACATTCGCATATGGCTGGAAGACACACAGTC	1915
Qy	500	AspGluIleGluLysAlaValLysLysAlaAsnAlaTyrAspPheIleMetLysLeuPro	519
Db	1916	GATGAGATTGAAAGCTGTCAAGAGACCAATGCCCTATCAGCTTATCATGAACATGCCCT	1975
Qy	520	AsnLysPheAspThrLeuValGlyLysArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539
Db	1976	CATAAATTTGACACCTTGGTTGGAGAGAGAGGGGCCAGTTGASTGGTGGGACAGACAGC	2035
Qy	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559

QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAAATCTTTGAGGAAGCACACATC 3235
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTGGAAATTACATTTTCCCTTACCAGCAATGATGATTTTCTCTATGCTGGATGTTTC 3295
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 3296 CGGTTTGGAGCCTACTTGGTGCAATAAATCATGAGCTTTGAGGATGTTCTGTAGTA 3355
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlnValSerSerPheAlaProAsp 999
Db 3356 TTTTTCAGCTGTTGTTTGGTGCCATGCGCGTGGGCAAGTCAGTTTCATTTCTCCTGAC 3415
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAGCCAAAATATCAGCAGCCACATCATCATCATTCGAAAAACCCCTTGG 3475
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTGACAGCTACAGCAGCGAAGCCCTAATCGCGAACACATTTGGAAGAAATGTCACATTT 3535
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGTATTCACATATCCACCCGACCGGACATCCCATGCTTCAGGAGCTGAGC 3595
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValIcLysSerSerGlyCysGlyLysSer 1079
Db 3596 CTGAGGTGAAGAAGGCCACAGCGCTGGCTGTGGGCGAGCTGGCTGTGGGAAGAGC 3655
QY 1080 ThrValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3656 ACAGTGTCTCAGCTCTCGGAGCGGTTCTACGACCCCTTGGCAGGGAAGTGTGCTGTGAT 3715
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 GGCNAGAAATNAAGCAGCTGAATGTTCAGTGGCTCCGAGCACACCTGGGCATCGTGCC 3775
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 CAGGAGCCCATCTGTTTGACTCGAGCATTTGCTGAGAACATTTGCCATGAGACAAACAGC 3835
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3836 CGGGTGTGTGCACAGGAAGAGATCGTGAGGCGCAGAAAGGAGCCCAATACATGCTTC 3895
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCAGTGCATCTGCTATAATAATAGCACCTAAGTAGGAGACAAAGAACTCAGCTCTCT 3955
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GGTGGCCAGAAACAACCATTTGCCATAGCTCGCTCGCTTTAGACAGCCCTCATATTTTG 4015
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGTGATGAAGCCAGCTCAGCTCTGGATACAGAAGTGAAGAGTTGTCCAAAGAGCC 4075
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGGACAAAGCCACAGAAAGCCGACCTGCATTTGTGATTGCTCACCGCCTGTCCACCATC 4135
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGAAATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGTCAAGGAGCATGGCAGCAT 4195
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGTGCTCAGTGTCCAGGCTGGAACA 4255

QY 1280 LysArg 1281
Db 4256 AAGCGC 4261
RESULT 6
US-08-784-649A-1
; Sequence 1, Application US/08784649A
; Patent No. 5830697
; GENERAL INFORMATION:
; APPLICANT: Sikic, Branimir I
; APPLICANT: Chen, Gang
; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
; TITLE OF INVENTION: CYCLOSPORIN MODULATION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/784,649A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: Reg.No. 5830697 36,677
; REFERENCE/DOCKET NUMBER: 06037/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-322-5070
; TELEFAX: 415-854-0875
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4264 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-784-649A-1
Alignment Scores:
Pred. No.: 0 Length: 4264
Score: 5815.50 Matches: 1162
Percent Similarity: 94.93% Conservative: 55
Best Local Similarity: 90.64% Mismatches: 60
Query Match: 89.83% Indels: 6
DB: 2 Gaps: 5
US-09-672-725C-23 (1-1281) x US-08-784-649A-1 (1-4264)
QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 143 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAAAGAACTTTTAAACTGAAC 202
QY 20 LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 203 AATAAAGTGAAGAAA---GATAAGAGGAAGAAACCACTGTCAGTGTATTTCAATG 259
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
Db 260 TTTTCGCTATTCAAAATG-CTTCACAAGTTGATATGTTGGTGGGAACCTTGGCTGCCATC 318
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 319 ATCCCATGGGCGGAGACTTCCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 378
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99

Db 379 GCAAAATGCAGGA---AATTTAGAACATCTGATGTCAACACATCACTAATAAGAGTATATC 435
 QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
 Db 436 AATGATACAGGGTCTTTCATGAAT---CTGGAGGAAGACATGACAGGTATGCCCTATTAT 492
 QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
 Db 493 TACAGTGAATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 552
 QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
 Db 553 CTGGCAGCTGGAAGACAAATACAAAAATAGAAAAACAGTTTTTCATGCTATATATGCGA 612
 QY 160 GlnGluIleGlyTyrPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
 Db 613 CAGGAGATAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 672
 QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
 Db 673 GATGCTCCAAAGATTAAAGAGGAATGGTGACAAAATTTGGAAGT---TTTCAGTCAATG 729
 QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTyrLysLeuThrLeuVal 219
 Db 730 GCAACATTTTCACTGGGTTTATAGTAGGATTTACACGTGGTGGTGGTGGTGGTGGTGGTGG 789
 QY 220 IleLeuAlaIleSerProValIleGlyLeuSerAlaAlaIleTyrAlaLysIleLeuSer 239
 Db 790 ATTTGGCCATCAGTCTGTTCTGGACCTGTCAGCTGCTGGTGGGCAAGATACTATCT 849
 QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
 Db 850 TCATTTACTGATAAAGAACTCTTAGCGTATGCCAAAGCTGGAGCAGTAGTGAAGAGGTC 909
 QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
 Db 910 TTGGCAGCAATTAGAACTGTGATTCATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 969
 QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
 Db 970 AACAAAATTTAGAAAGCTTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCAAATATT 1029
 QY 300 SerIleGlyAlaAlaPheLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrTyrGly 319
 Db 1030 TCTATAGTGCTGCTTCTCTGCTGATCTATGCACTTATGCTCTGGCCTTCTGGTATGGG 1089
 QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
 Db 1090 ACCACCTTGGTCTCTCAGGGGATATTTCTATTGGACAAAGTACTCACTGTATTTCTTCT 1149
 QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
 Db 1150 GTATTAAATTTGGGCTTTTGGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAAAT 1209
 QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
 Db 1210 GCAAGAGGACAGCTTATGAAATCTCAAGATAATTGATAATTAAGCAAGTATTGACAGC 1269
 QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
 Db 1270 TATTGGAAGAGTGGGCAACACAGATAATTAAAGGAAATTTGGAATTCAGAAATGTT 1329
 QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
 Db 1330 CACTTCAGTTACCCATCTCGAAAGAAAGTTAAGATCTTGAAGGGCTGAACCTGAAGGTTG 1389
 QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
 Db 1390 CAGAGTGGCAGAGCGGTGGCCCTGGTTGGAAACAGTGGCTGGTGGGAGGACCAACAGTCT 1449
 QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
 Db 1450 CAGCTGATGCAGAGGCTCTATGACCCACACAGAGGGATGGTCACTGTTGATGACAGGAT 1509

Db 1450 CAGCTGATGCAGAGGCTCTATGACCCACACAGAGGGATGGTCACTGTTGATGACAGGAT 1509
 QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
 Db 1510 ATTAGGACCAATAATGTAAGGTTTCTACGGAAATCATTTGTTGTTGTTGTTGTTGTTGTTGTT 1569
 QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
 Db 1570 GTATTGTTTGGCCACCATAGCTGAAACATTCGGTATGGCCGTGAAATGTCACCATG 1629
 QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
 Db 1630 GATGAGATTGAGAAAGCTGTCAGGAAGCAATGCCATGACTTTATCATGAACCTGCCT 1689
 QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
 Db 1690 CATAAATTTGACACCCCTGGTTGGAGAGAGAGGGGCCAGTTGAGTGGTGGGCAAGACGAG 1749
 QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAla 559
 Db 1750 AGGATGCCCTTCACGTGCCCTGGTTCGCAACCCCAAGATCCTCTGCTGGATGAGGC 1809
 QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
 Db 1810 ACCTCAGCCTTGACACAGAAACCGAAGCAGTGTTCAGGTGGCTCTGGATGAAGCCACA 1869
 QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
 Db 1870 AAAGGTGGACCACTTGTGATGCTCATCGTTTCTACAGTTGCTAATGCTGACGTC 1929
 QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
 Db 1930 ATCCCTGGTTTCGATGATGAGTCATTTGGAGAAAGAAATCATGATGAATCTCATGAAA 1989
 QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
 Db 1990 GAGAAAGGCATTTTACTCAAACTTGTCACAAATGCACAGCAGCAAGAAATGAATGAATTA 2049
 QY 640 GluAsnAlaThrGlyLysSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
 Db 2050 GAAATATGCACTGATGAATCCAAAGTGAATGATGCTTGGAAATGCTTCAATGAT 2109
 QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
 Db 2110 TCAGATGACAGCTATATAAGAAAAGATCACTCTGATGAGGTGCTCCGTGGATCAACAGCC 2169
 QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
 Db 2170 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGTATACCTCCAGTTTCCITT 2229
 QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCys 719
 Db 2230 TGGAGATTATGAAGCTAAATTTAACTGAATGCTTATTTGTTGTTGTTGTTGTTGTTGTT 2289
 QY 720 AlaIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
 Db 2290 GCCATTATAATGGAGGCTCGAACAGCATTTGCAATAATATTTTCAAGAAATTAAGG 2349
 QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
 Db 2350 GTTTTACAGAATTTGATGATCCTCGAAACAAACACAGCAATAGTAACCTGTTTCACTA 2409
 QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
 Db 2410 TTGTTTCTAGCCCTTGGAAATTTATTTCTTTATTTATTTTCTCTCAGGGTTTCATATT 2469
 QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
 Db 2470 GCAAAAGCTGGAGAGATCTCCCAAGCGCTCCGATCATGTTTCCGATCCATGCTC 2529
 QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
 Db 2530 AGACAGGATGTGAGTTGTTGATCACCCCTAAAAACACCACCTGGAGCATTCACCTACCAGG 2589

QY 820 LeuAlaAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
 Db 2590 CTGGCAATGATGCTCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTAATACC 2649
 QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrglyTrpGlnLeu 859
 Db 2650 CAGAAATAGCAAACTCTGGGACAGGAATAATATATCTCTCATCTATGTTGGCAACTA 2709
 QY 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
 Db 2710 ACACCTGTTACTCTTAGCAATTTGACCATCATTTGCAATAGCAGGAGTTGTTGAAATGAAA 2769
 QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGluValAlaGlyLysIleAla 899
 Db 2770 ATCTTGTCTGGACACACACTGAAAGATAAGAAAGACTAGAAGGTCTGGGAAGATCGCT 2829
 QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
 Db 2830 ACTGAAGCAATAGAAACTTCGACCGTGTCTTCTTCTGACTCAGGAGCAGAAAGTTTGA 2889
 QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
 Db 2890 CATATGTATGCTCAGAGTTTGCAGTACCATACAGAAACTCTTTGAGAAAGCACACATC 2949
 QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
 Db 2950 TTTGGAAATACATTTCTTCCACCGACGAGCATATATTTTCTTCTGATGCTGGAATTTTC 3009
 QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
 Db 3010 CGGTTTGGAGCCCTACTTGTGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 3069
 QY 980 PheSerAlaIleValPheGlyAlaMetAlaValAlaGlyGlnValSerPheAlaProAsp 999
 Db 3070 TTTTACGCTGTGCTTGTGTGTGTCATGGCGTGGGCAAGTCAGTTTCATTTGCTCTGAC 3129
 QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
 Db 3130 TATGCCAAAGCCAAATATCAGCAGCCCATCATCATCATCATCATCATCATCATCATCAT 3189
 QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
 Db 3190 ATTGACAGCTACAGCAGGAGGCGCTAATGCCGAACACATTTGGAAGAAATGTCACATTT 3249
 QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
 Db 3250 GGTGAAGTTGTTTAACTATCCACCCGACCGCATCCCATGCTTTCAGGAGCTGAGC 3309
 QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
 Db 3310 CTGGAGGTGAAGAGGCGCCAGACGCTGGCTCTGTTGGCAGCATGCTGCTGGAAGAGC 3369
 QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
 Db 3370 ACAGTGTCCAGCTCTCTGAGCGGTTCTACGACCCCTTGGCAGGAAAGTCTGCTGTAT 3429
 QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
 Db 3430 GGCAAGAAATAAAGCCACTGAATGTTCACTGGCTCGAGACACACCTGGGCACTGCTGCC 3489
 QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
 Db 3490 CAGGAGCCCATCTGTTTACTGTCAGCATTTGCTGAGAACAATGCTATGAGAGACAACAGC 3549
 QY 1140 ArgValValSerHisGluGluIleMetGlnAlaLysLysGluAlaAsnIleHisHisPhe 1159
 Db 3550 CGGTGTGTCTCAGAGGAAGATTTGTGAGGCGCAGCAAGAGGCGCAATACATGCGCTTC 3609
 QY 1160 IleGluThrLeuProGluGluAspTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
 Db 3610 ATCAGTCTACCTGCTTAATAATATAGCACTAAAGTAGAGACAAAGAACTCAGCTCTCT 3669

QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
 Db 3670 GGTGGCAGAAACAACGCAATGCCATAGCTCGTCCCTTGTAGACAGCCATCATATTTTG 3729
 QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
 Db 3730 CTTTGTGATGAAGCCAGCTCAGCTCTGATACAGAAAGTGAAGAGTTGTCRAAGAGCC 3789
 QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
 Db 3790 CTGGACAAGCCAGAGAAGCGCCACCTGCTGATTTGCTACCCGCTGTCCACCATC 3849
 QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
 Db 3850 CAGAATGCAGACTTAATAGTGTGTTTTCAGAAATGCCAGAGTCAAGAGCATGGCAGCAT 3909
 QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
 Db 3910 CAGCAGCTGCTGGCAGACAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGACA 3969
 QY 1280 LysArg 1281
 Db 3970 AAGCGC 3975
 RESULT 7
 US-08-784-649A-5
 ; Sequence 5, Application US/08784649A
 ; Patent No. 5830697
 ; GENERAL INFORMATION:
 ; APPLICANT: Sikic, Branimir I
 ; APPLICANT: Chen, Gang
 ; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
 ; TITLE OF INVENTION: CYCLOSPORIN MODULATION
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 2200 Sand Hill Road
 ; CITY: Menlo Park
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/784,649A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sherwood, Pamela J
 ; REGISTRATION NUMBER: Reg.No. 5830697 36,677
 ; REFERENCE/DOCKET NUMBER: 06037/007001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-322-5070
 ; TELEFAX: 415-854-0875
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4264 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-08-784-649A-5
 Alignment Scores:
 Pred. No.: 0 Length: 4264
 Score: 5815.50 Matches: 1162
 Percent Similarity: 94.93% Conservative: 55
 Best Local Similarity: 90.64% Mismatches: 60
 Query Match: 89.83% Indels: 6
 Gaps: 2
 DB:

US-09-672-725C-23 (1-1281) x US-08-784-649A-5 (1-4264)

Qy 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheThrLysMetGly 19
Db 143 ATGGATCTTGAAGGGACCGCAATGGAGGACGAAGAAGAAGAACTTTTAAACTGAAC 202
Qy 20 LysLysSerLysLysAsnGluLysLysLysProThrValSerThrPheAlaMet 39
Db 203 AATAAAAGTGAATAA---GATAAGAAGAAAGAAACCAACTGTCAGTGTATTTTCAATG 259
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 260 TTTCGCTATTCAAAATG-CITGACAAGTGTATATGCTGGTGGGAACCTTTGGCTGCCATC 318
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 319 ATCCATGGGGTGGACTTCCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 378
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 379 GCAAATGCAGGA---AATTTAGAAGATCTGATGTCAAAACATCACTAATAGAAGTGATATC 435
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTrpAlaTyrTyr 119
Db 436 AATGATACAGGGTCTTCATGAAT--CTGGAGGAAGACATGACCAAGTATGCCATATAT 492
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 493 TACAGTGAATTTGGTGTGGGTGCTGCTGCTTACATTCAGGTTTCATTTTGGTGC 552
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 553 CTGCAGCTGGAAGACAAATACAAAATTAGAAAAAGTTTTCATGCTATAATGCGA 612
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 613 CAGGAGATAGCTGGTTGATGTCACAGATGTTGGGAGCTTAACCCGACITACAGAT 672
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 673 GATGCTCCAAAGATTAATGAAGGAATTTGGTGACAAAAATTTGGAATG---TTCCAGTCAATG 729
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 730 GCAACATTTTTCAGTGGGTTTATAGTAGGATTTACACGTGGTGGGAAGCTAACCCCTTGTG 789
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 790 ATTTTGGCCATCAGTCCCTGTTCTTGGACTGTACAGCTGCTGTCTGGCAAGATACTATCT 849
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
Db 850 TCATTTACTGATAAAGAACTCTTAGCGTATGCCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 909
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyr 279
Db 910 TTGCAGCAATTAGAACTGTGATTGCAATTTGGAGGACAAAGAAAGAACTTTGAAGGTAC 969
Qy 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 970 AACAAAAATTTAGAAGAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCAAATTT 1029
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 1030 TCTATAGGTGCTGCTTCTGCTGATCTATGCACTTTATGCTCTGCGCTTCGTGTATGGG 1089
Qy 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 1090 ACCACCTTGGTCTCTCAGGGGAATATCTATTGGACAAGTACTCACTGTATTTCTTCT 1149
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
|||||

Db 1150 GTATTAATTTGGGCTTTTAGTGTGGACAGGCATCTCCAAGCATTTGAAGCATTTGCCAAAT 1209
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1210 GCAAGAGGAGCAGCTTATGAAATCTTTCAAGATAATTTGATAAAGCCAAGTATTGACAGC 1269
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1270 TATTCACAGAGTGGGCACAAACAGATAATATTAGGGAAATTTGGAATTCAGAAATGTT 1329
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1330 CACTTCAGTTACCCATCTCGAAAGAAGTTAAGATCTTGAAGGGCTGAACCTGAAGGTG 1389
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1390 CAGAGTGGGCAGACGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGACACAACAGTC 1449
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1450 CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGATGTCAGTCTTGATGCACAGGAT 1509
Qy 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1510 ATTAGGACCATAAATGTAAGGTTTCTACGGGAAATCATTTGGTGTGTCAGTGCAGGAACCT 1569
Qy 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1570 GTATTTTGGCCACCACGATAGCTGAAACATTCGCTATGGCCGTGAAATGTCCACCATG 1629
Qy 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1630 GATGAGATTTGAGAAAGCTGTCAAGGAAGCCATGCCATGACTTTATCATGAACCTGCCT 1689
Qy 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1690 CATAAATTTGACACCCCTGGTTGGAGAGAGAGGGGCCAGTTGAGTGGTGGGAGAGCAG 1749
Qy 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 1750 AGGATGCCATTCGACGTCCTGGTTGCGAACCCCAAGATCCTCTGCTGATGAGGCC 1809
Qy 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 1810 ACCTCAGCCTTGACACAGAAAGCAGGAGCAGTGGTTTCAGTGGCTCTGATGAAGCCACA 1869
Qy 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 1870 AAAGGTCGGACCACTTGTGATAGCTCATCGTTTCTCTACAGTTCTGTAATGTCGACGTC 1929
Qy 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 1930 ATCGCTGGTTTCGATGATGAGTCATTTGGAGAAAGAAATCATGATGAATCATGATAA 1989
Qy 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 1990 GAGAAAGGCATTTACTTCAAACTGTCAATGCACACAGCAGGAATAAGAGTTGAATTA 2049
Qy 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2050 GAAAATGCAGCTGATGAATCCAAAAGTGAATTTGATGCGCTTGGAAATGTCTTCAATGAT 2109
Qy 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2110 TCAAGATCCAGTCTAATAAGAAAAGATCAACTCGTAGAGAGTGTCCGTCGGATCACAAGCC 2169
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2170 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCTT 2229
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2230 TGCAGGATTATGAAGCTAAATTTAACTGAATGGCCCTATTATTTGTTGTTGTTGTTGTTGTT 2289

1	MetAspProGluGlyArgLysGlySerAla---	GluLysAsn	PheTrpLysMet	Gly19
425	ATGGCTATGAAGGGAGCCGCAATGGAGGAGCAAGAAAGAACTTTTTTAAACTTGAC	484		
20	LysLysSerLysLysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMet	39		
485	AATAAAGATGAAAA---GATAAGAAGAAACCAACTGTCAGTATTTTCAATG	541		
40	PheArgTrpSerAsnTrpLeuAspArgLeuTrpMetLeuValGlyThrMetAlaAlaIle	59		
542	TTTCCTATTCAATTCGCTTGACAGTGTATATGGTGGGAACTTTTGGCTGGCCATC	601		
60	IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe	79		
602	ATCCATGGGGTGGACTTCCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT	661		
80	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr	99		
662	GCAATATGACAGA---AATTTAGAAGATCTCATGTCCAAACATCATATAAGATGATATC	718		
100	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrrAlaTyrr	119		
719	AATGATACAGGGTCTTCATGAAT---CTGGAGGAGAACATGACACAGTATGCTATTAT	775		
120	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrrIleGlnValSerPheTrpCys	139		
776	TACAGTGGAAATTGGTGTGGGGTGTGGTGTCTTACATTACAGTTTCATTTTGGTGC	835		
140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159		
836	CTGGCAGCTGGAAGACAAAATACACAAAATAGAAAACAGTTTTTTCATGCTATAAATGCCA	895		
160	GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179		
896	CAGAGATAGCTGGTTTGAATGACAGATGTGGGACGTTTAACCCGACCTTACAGAT	955		
180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle	199		
956	GATGCTCTAAGATTATGAAGCTTATGGTGACAAAATGGAAATGTTCTTTTCAAGTCAATG	101		
200	AlaThrPheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219		
1016	GCAACATTTTTCATCGGTTTATAGTAGGATTTACACGTGGTTGGAAGCTAACCCCTTG	1071		
220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleThrPalLysIleLeuSer	239		
1076	ATTTTGGCCATCATGCTCTTCTGGACTGTGCTGTGGCAAGATPACTATCT	1131		

||||| 2216 ATCGCTGGTTGATGATGGAGTCAATCTGGAGAAAGCAATCATGATGAACATCATGAAA 2275
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 2276 GAGAAAGGCAATTTACTTCAAACTTGTCAATGTCAGACAGCAGGAAATGAAGTTGAATTA 2335
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2336 GAAATATGACGTGATGAATCCAAAAGTGAATTTGATGCTTGGAAATGCTCTCAATGAT 2395
QY 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2396 TCAAGATCAGCTTAATAAGAAAAGATCAACTCGTAGAGTGTCCGTGGATCACAACCC 2455
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2456 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGTATATACCTCCAGTTTCCCTT 2515
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2516 TGGAGGATTAAGCTAAATTAATTAAGTGAATGCGCTTATTTTGTGTTGGTGTATTTGT 2575
QY 720 AlaIleAsnGlyLeuGlnProAlaPheSerIlePheSerArgIleIleGly 739
Db 2576 GCCATTATAATGGAGGCTGCAACACGACATTTGCAATAATATTTTCAAGATTAAGGG 2635
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 GTTTTACAGAATGATGATCCTGAAACAAAACGACAGAACTACTAACTTTGTTTTCACTA 2695
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2696 TTGTTTCTAGCCCTTGGAAATTAATTTCTTTTATTTACATTTTCTTCACGGTTTTCACAT 2755
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 GGCAAAAGCTGGAGAGATCCTCACAAGCGCTCCGATACATGTTTCCGATCCATGCTC 2815
QY 800 ArgGlnAspValSerThrPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2816 AGACAGATGTGAGTTGGTTGATGACCTTAAACACACACCTGGAGCATGTACTACCAG 2875
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2876 CTCGCCAATGATCTCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTCTGCTGAATTACC 2935
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2936 CAGAATATAGCAATTTCTGGGACAGGAATAATATATCTTCATCTATGTTGGCAACTA 2995
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2996 ACATGTTACTCTTAGCAATTTGACCCATTCATTCGAATAGCAGGAGTTGGAATGAAA 3055
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 3056 ATGTTGCTGACACAGCACTGAAAGATAAGAAAGAACTAGAAGTGCTGGGAAGATCGCT 3115
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 3116 ACTGAAGCAATAGAAACTTCCGAACCGTTGTTTCTTCTACTCAGGACAGAGAGTTTGA 3175
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAAATCTTTGAGGAAGCACACATC 3235
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTTGAATTAATTTCTTCCACCCAGGCAATGATGATTTTTCCTATGCTGATGTTTC 3295
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
|||||

Db 3296 CGGTTTGGAGCCCTACTTGGTGGCACATAAACTAATGAGCTTTGAGGATGTTCTGTAGTA 3355
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3356 TTTTACGCTGTTGCTTTGCTGCGATGGCGGTGGCAAGTGAAGTTCAATTTGCTCCTGAC 3415
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAAGCAAAATATCAGCAGCCACATCATCATGATCATTTGAAAACCCCTTTG 3475
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTACAGCTACACGAGCGGAAAGCTTAATGCCGAACACATTTGGAAGAAATGTACATTT 3535
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGATTAACATATCCACCCGACGACATCCAGTCTCTTCAGGACGTGAGC 3595
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3596 CTGGAGGTGAAGAGGCCAGACGCTGCTGCTGGCGACAGTGGCTGTGGGAAGAGC 3655
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3656 ACATGCTGACGCTCTCTGGAGCGGTTCTACACCCCTTGGCAGGAAAGTGTCTGTAT 3715
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 GGCAAAAGAAATAAGCGACTGAATGTTTCACTGGCTCGGACACACCTTGGGCATCGTGC 3775
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 CAGGAGCCCATCTCTGTTGATGCTGAGCAATGCTGAGAACATTTGCTATGGACACACAGC 3835
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3836 CGGTGTTGTCACAGGAAGAGATCGTGAGGCGCAAAAGGCGCAACATACATGCTTTC 3895
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCGAGTCACTGCTTAATAATATAGCACTAAAGTAGGACACAAAGGAAGTCACTCTCT 3955
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GGTGCCAGAAACAAACGATTTGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4015
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGTGATGAAGCCAGCTCAGCTCTGATACAGAAAGTGAAGAGTTGTCCAAAGAGCC 4075
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGGACAAACCCAGAGAAGCGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4135
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGATGTCAGACTTAATAGTGTGTTTTCAGAAATGCGAGAGTCAAGCAGCATGGCAGCAT 4195
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGCTGGCACAGAAAGGCACTATTTTCAATGGTCACTGTCCAGCTCGGAACA 4255
QY 1280 LysArg 1281
Db 4256 AAGCGC 4261

RESULT 9
US-09-120-513-1
; Sequence 1, Application US/09120513
; Patent No. 6025160
; GENERAL INFORMATION:
; APPLICANT: Brun, Kimberly
; APPLICANT: Chenery, Richard

APPLICANT: Ellens, Harma
APPLICANT: Field, John
APPLICANT: Yue, Lin
TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE
TITLE OF INVENTION: SEQUENCES ENCODING RAT MDR1B2 AND
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY:
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/120,513
FILING DATE: 22-JUL-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: King, William T
REGISTRATION NUMBER: 30,954
REFERENCE/DOCKET NUMBER: GP50008
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5015
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4233 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-120-513-1

Alignment Scores:
Pred. No.: 0 Length: 4233
Score: 5291.50 Matches: 1030
Percent Similarity: 90.48% Conservative: 130
Best Local Similarity: 80.34% Mismatches: 113
Query Match: 81.73% Indels: 9
DB: 3 Gaps: 5

US-09-672-725c-23 (1-1281) x US-09-120-513-1 (1-4233)

```
QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPhetrpLysMetGlyLys 20
Db 26 ATGGAGTTTGAAGAGGCGCTTAACGGAAACAGCAGCAAGAACTCTCAAGATGGGCMAA 85
QY 21 LysSerLysLysAsnGluLysLysGluLysProPhrValSerThrPheAlaMetPhe 40
Db 86 AAGAGTAAAG-----GAGAGGAGAGAAACCTGCTTGGCATATTGGGATGTTT 139
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 60
Db 140 CGCTATGCAGATTGGCTTGACAAGCTGTGCATGGCTCTGGGAACCTCTCGCTGCTATCATC 199
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
Db 200 CACGGAAACCTGCTTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 259
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 260 CAAGCA-----GAGACCCCGCATCTGCCAGCGTTACTAATCAAGTGAATCAAC 310
QY 101 AsnThrGlnHisPheIle----AsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 311 AGTACACAGACCGTCTGAGCGACAGCAGTCTGGAGGAGGACATGCCCATGTGACGCTACTAT 370
```

```
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 371 TACACGGGCATGGTGGCGGTGCTCATCGTTGCCGTACATCCAGGTTTCACTTTGGTGC 430
QY 140 LeuAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 431 CTGGCAGCTGGGAGACAATAACACAAGATTAGCAGAAGTTTTCATGCCATCATGAAT 490
QY 160 GlnGluIleGlyTrpPheAspValHisAspValIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db 491 CAGGAGATAGGCTGGTTTGACGTGAATGACGCTGGGAGCTCAACACCCGGCTCACAGAT 550
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db 551 GACGCTCCAAAATAATGACGGAATTGGTGACAACTTGAATGTTCTTTTCAGTCCATA 610
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrVal 219
Db 611 ACACATTTTCAGCCCGTGTATTAATAGGATTTATAGTGGTTGGAGGCTAACCCCTTGT 670
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 671 ATTTTGGCGTCAGCCCTCTTATTGGTGTGTCATCTGCCATGGGCAAGGTACTGACT 730
QY 240 SerPheThrAspLysGluLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
Db 731 TCATTTTACTAATAAAGGAACCTCCAGGCTTATGCCAAAGCTGGAGCAGTTCGCCAAGAAGTC 790
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 791 TTAGCAGCCATCAGAACTGTGATTGGTTGGAGGACAAAGAAAGAACTTGAAGGTAC 850
QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 851 AATAAAAAATTTAGAAGAAGCTAAAGAGTTGGCATAAAGAAAGCCATCACGCCCAACATT 910
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 911 TCCATAGGTATTGCCCTACCTGTGGTCTATGGTCTTATGCATGGCATCTCGTATGGG 970
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 971 ACCTCCTTGGTCTCTCAAAATGAATATCTATTGGCAAGTCTTACCGCTCTTCTCTCT 1030
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1031 ATTTTATTGGGACCTTTCAGTATTGGACATTTAGCCCAACATAGAGCCCTTTCGCAAT 1090
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1091 GCAAGAGGGGAGCCTATGAAATCTTCAAGATAATTGATAATGAGCCAAGCATTGACAGC 1150
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1151 TTCTCAACCAAGGACACAAACACAGACATGATAATGGGAAATTTGGAATTTAAATAATGTT 1210
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1211 TACTTCAACTACCCATCACGAAGTGAAGTTAAGATCTTGAAGGGCCTCAACCTGAAGGTG 1270
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1271 AAGAGCGGCGACAGCGTAGCCCTGGTTGGCAACAGTGGCTGTGGGAAAAGCAACAATGTC 1330
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1331 CAGCTGCTGCAGAGGCTCTACGACCCCATAGAGGCGGAGGTGAGTATCGACGACAGGAC 1390
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValIleSerGlnGluPro 479
Db 1391 ATCAGGACCATCAATGTGAGGTATCTGCGGGAAATCATTTGGGGTGGTGGTGGTGGTGGT 1450
```

QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArGfYrGlyArgGluAsnValThrMet 499
 |||||
 Db 1451 GTGCTGTTTCCACCACGATTGCCGAAACATTCGTATGGCCGAGAAACGTCACCATG 1510
 QY 500 AspGluIleGluYsAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
 |||||
 Db 1511 GATGAGATAGAAAAGCTGTCAAGGAAGCAATGCTATGACTTCATCATGAAACTGCCC 1570
 QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
 |||||
 Db 1571 CACAAATTTACACCTGTTGGTGAGAGAGGGCGAGCTGAGTGGGGGACAGAAACAG 1630
 QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAla 559
 |||||
 Db 1631 AGGATGCCATTGCCGGGCCCTGTGTCGCAACCCCAAGATCCTTTTGTGGATGAGGCC 1690
 QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
 |||||
 Db 1691 ACGTACAGCTTGACACAGAAAGCGGCGTGGTTCAGGGCGCTCTGGATAAGGCTAGA 1750
 QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
 |||||
 Db 1751 GAAGCGCGGACACCATTTGTGATGCTCACCGCTTGTCTACAGTGGCAATGCTACGCTC 1810
 QY 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
 |||||
 Db 1811 ATTGCTGGTGGTGTGCTGCTATTGTGGAGCAAGGAATCATGAAGAGCTCATGAAA 1870
 QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
 |||||
 Db 1871 GAGAAGGGCATTTACTTCAAACTTGTCTATGACACAGACTAGAGGAATGAATGAACCA 1930
 QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
 |||||
 Db 1931 GGAATAATGCTATGAAATCCCAAGTACACTGGTGGCTCTGAGTTGACTTCAGAAAA 1990
 QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
 |||||
 Db 1991 TCAAAATCTCCTTTAATA---AGGAGATCAATTCCGAGAAGTATCCACAGAACAGAC 2047
 QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
 |||||
 Db 2048 CAGGAGAGAAGCTTAGTTCGAAGAGGATGGATGAAGATGTGCTATGGTTTCCCTT 2107
 QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
 |||||
 Db 2108 TGGCAGATCCTAAAGCTAAATATTAGTGAATGCGCCCTATTAGTTGTGGGTCTACTTGT 2167
 QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
 |||||
 Db 2168 GCTGTTAATAATGGGTGCATACACAGGTGTTGGCATAGTGTTCAAAGATTGTAGGG 2227
 QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
 |||||
 Db 2228 GTTTTTTCAAGACGACGACCCATGAACCAACAAACAGGAATGTAACTGTGTTCCCTT 2287
 QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
 |||||
 Db 2288 CTCTTCTGTCATGGGAATGATTCTTTTGTACGTACTTCTTTCAAGGCTTCATATT 2347
 QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
 |||||
 Db 2348 GCGAAAGCTGGAGAGATCCTCAACAGCGACTCCGATACATGGTCTTCAAAATCCATGCTG 2407
 QY 800 ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
 |||||
 Db 2408 CGACAGATATAGCTGGTTGTATGACCATTAACCAACACACTGGCTCGCTGACTACCAGG 2467
 QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
 |||||
 Db 2468 CTCGCTAGTCACGCTTCTAATGTTAAAGGGCTATGGGCTCCAGGCTTGTCTAGTTACC 2527
 QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
 |||||

Db 2528 CAGAAATGTAGCAACCTTGGCAGAGGAATTATCTTATCCCTTAGTCTATGGCTGGCAGCTT 2587
 QY 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
 |||||
 Db 2588 ACACCTTTTACTTGTAGTAATATACCACTCATTTCTCTGGTGGAAATATTGAATGAAA 2647
 QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 899
 |||||
 Db 2648 CTGTTGTCTGTCTCAAGCTTTGAAGGACAAAGAGAGTAGAGATCTCTGGGAAGATCGCT 2707
 QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
 |||||
 Db 2708 ACAGAAACAATTGAAACTTCCGCACCTGTGCTCTTTGACTAGGAGAGCAAGATTGAA 2767
 QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
 |||||
 Db 2768 ACTATGTATGCCAGAGCTTGCAGATACCATACAGAAATGCTTTGAAGAAAGCACAGTC 2827
 QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
 |||||
 Db 2828 TTTGGGATCACCTTCGCTTCACCCAGGCCATGATTTATTTTCTATGCTGTGTTTC 2887
 QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
 |||||
 Db 2888 CGGTTTCGGTGGCTACTTGGTGGCAGAGACTCATGACGTTTGAAATGTTATGTTGGTA 2947
 QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
 |||||
 Db 2948 TTTTCTGCTGTGCTTGGTGCCATGCGCAGCGAATACCACTTCATTCGCTCCTGAC 3007
 QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
 |||||
 Db 3008 TACGGAAGGCCAAAGCTTCAGCATCCACATCATAGGATCATTTGAGAAATCCCCGAG 3067
 QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
 |||||
 Db 3068 ATTGACAGCTACACGCGAGGGCTTGAAGCTAATTTGGTTAGAAAGAAATGTGAATTT 3127
 QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
 |||||
 Db 3128 AATGGAGTCACTGTTCAACTATCCACCCGACCCCAACATCCAGTCTTCAGGGACTGAGC 3187
 QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysLysSer 1079
 |||||
 Db 3188 TTCAGGTGAAGAAGGGGCAACGCTTCGGCTGTGGCAGCAGTGGCTCGGGAAGAGT 3247
 QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
 |||||
 Db 3248 ACAGTGTCCAGCTGCTCGAGCGCTTCTACAACCCCATGGCTGGACAGTGTTCCTAGAT 3307
 QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
 |||||
 Db 3308 GCGAAAGAAATAAACAACTCAAGTCCAGTGGCTCGCGGCCCACTTGGGCATTTGTCTCC 3367
 QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
 |||||
 Db 3368 CAGAGGCCATCTCTGTTGACTGCAGCATCCCGAGAACATCGCCTACGGAGACACAGC 3427
 QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
 |||||
 Db 3428 CGTGTGCTCATGAGGAGATCGTGGGGCCCGCAGGGAGGCCAACATCCACCACTTC 3487
 QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
 |||||
 Db 3488 ATCGACTCACTGCTGAGAAATACAAACAGAGTGGGAGACAAAGGAGCTCAGCTCTCG 3547
 QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
 |||||
 Db 3548 GCGGGGAGAGACGCGCATCGCCATCGCGGCCCTCTGTCAGACAGACCTCATCTTA 3607
 QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
 |||||

Db 3608 CTTCTGGATGAAGCGACATCATGCTCTGGATACGAGAGTGAAGAGGTGCTCCAGGAAGCG 3667
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3668 CTGGACAAGCAGGGAAGCGCCACCTGCTGATGTGATCGCGCCGCTGTCCACCATC 3727
QY 1240 GlnAsnAlaAspLeuIleValIlePheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3728 CAGAAGCAGACTTGTATGCTGATTCAGAACGCCAGGTCAAGGAGCAGCGCCACCAC 3787
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3788 CAGCAGCTGCTGGCCAGAAAGGCATCTATTCTCGATG-----GTTCAAGCTGGAGCA 3841
QY 1280 LysArg 1281
Db 3842 AAGCCG 3847
RESULT 10
US-09-450-105-1
; Sequence 1, Application US/09450105
; Patent No. 6169166
; GENERAL INFORMATION:
; APPLICANT: Kimberly Anne Brun
; APPLICANT: Richard James Chenery
; APPLICANT: Harma Ellens
; APPLICANT: John Anthony Feild
; APPLICANT: Lin Yue
; TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE SEQUENCES
; FILE OF INVENTION: ENCODING RAT MDR1B2 AND SCREENING METHODS THEREOF
; FILE REFERENCE: GP-50008-D1
; CURRENT APPLICATION NUMBER: US/09/450,105
; CURRENT FILING DATE: 1998-11-29
; EARLIER APPLICATION NUMBER: 09/120,513
; EARLIER FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4233
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-450-105-1
Alignment Scores:
Pred. No.: 0 Length: 4233
Score: 5291.50 Matches: 1030
Percent Similarity: 90.48% Conservative: 130
Best Local Similarity: 80.34% Mismatches: 113
Query Match: 81.73% Indels: 9
DB: 4 Gaps: 5
US-09-672-725C-23 (1-1281) x US-09-450-105-1 (1-4233)
QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 26 ATGGAGTTGAAGAGGCGCTTAACGGAACAGCAGCAAGAACTTCTCAAAAGATGGCAAA 85
QY 21 LysSerLysLysAsnGluLysGlyLysLysProThrValSerThrPheAlaMetPhe 40
Db 86 AAGAGTAAAAAG-----GAGAAGGAGAAGAAACCTGCTGTGGCATATTTCGGGATGTTT 139
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 60
Db 140 CGCTATGCGAGTGGCTTGACAAGCTGTCATGCTCGGAACTCTCGCTGCTATCATC 199
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
Db 200 CACGGAACCCCTGCTCCCTCTGATGCTGTGTGGATACATACAGATAGTTTACC 259
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 260 CAAGCA-----GAGACCCGCATCTGCCGAGCGTTACTAATCAAGTGAATCAAC 310

QY 101 AsnThrGlnHisPheIle---AsnHisLeuGluGluMetThrThrThrTyrAlaTyrTyr 119
Db 311 AGTACACAGACCGCTCAGCAGACAGCTCTGGAGAGAGACATGCCATGTACGCCCTACTAT 370
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 371 TACAGCGGCATGTGTCGGTGTGCTCATGCTTCCCTACATCCAGGTTTCACTTTGGTGC 430
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 431 CTGGCAGCTGGGAGACAAATACACAGATPAGCAGAAGTTTTTCCATGCCATCAATGAAT 490
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 491 CAGGAGATAGCTGGTTTGACGTGAATGACGCTGGGAGGCTCAACACCGGCTCACAGAT 550
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db 551 GAGCTCTCCAAAATTAATGACGGAATTTGGTACAAAATTGGAATGTCTTTCAGTCCATA 610
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 611 ACGACATTTTCAGCCGGTTTTTAATAGGATTTAATAGTGGTTGGAAGCTAAACCTTGTA 670
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 671 ATTTTGGCGCTCAGCCCTCTTATTTGGTGTCTATCTGCCATGTGGCAAGAGTACTGACT 730
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 731 TCATTACTAATAAGGAACCTCCAGGCTTATCGAAAAGCTGGAGCAGTTGCCGAAGAAGTC 790
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 791 TTAGCAGCCATCAAACTGTGATTGCGTTTGGAGGACAAAGAAAGAACTTGAAGGATAC 850
QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 851 AATAAAATTTAGAGAAGCTAAAGAGTTGGCATTAAGAAAGCCATCAGGGCCACATTT 910
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 911 TCCATAGGTATTGCTACCTACCTGTTGGTCTATGCTCTTATGCACTGCAATTTCTGGTATGG 970
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 971 ACCTCTTGGTCTCTCAAAATGAATATTCTATTGGACAAAGTGTCTTACCGTCTTCTCTCT 1030
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1031 ATTTTATGGGAGCTTTCAGTATTGGACATTTAGCCCCCAACATAGAGCTTTGGCAAT 1090
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1091 GCAAGAGGGCAGCCTATGAAATCTTCAAGATAATTGATAATGAGCAAGCATTGACAGC 1150
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1151 TTCTCAACCAAGGACACAAACAGACAGTATTAATGGGAAATTTGGAATTTAAAAATGTT 1210
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1211 TACTTCACTACCCTACAGAACTGAAGTTAAGTATCTTGAAGGCCCTCAACCTGAAGGTG 1270
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1271 AAGAGCGGCGAGCAGGTAGCCCTGGTTGGCAACAGTGGCTGTGGGAAAAAGCACAACTGTC 1330
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1331 CAGCTGTGCAGAGGCTCTACGACCCCATAGAGGGCGAGGTAGTATCATCGACGACAGGAC 1390
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479

|||||
Db 1391 ATCAGGACCATCAATGTGAGGTATCTGCGGGAATCATTTGGGGTGGTGAGTCAGGAACCC 1450
Qy 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1451 GTGCTGTTGGCCACCAAGATGCCGAAACATTCGCTATGGCCGAGAAACGCTCACCATG 1510
Qy 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1511 GATGAGATAGAGAAAGCTGTCAAGGAAGCCAAATGCCATATGACTTCATCATGAAACTGCC 1570
Qy 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1571 CACAAATTTAACACCCCTGTTGGTGAGAGAGGGCGCAGCTGAGTGGGGGACAGAAACAG 1630
Qy 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAla 559
Db 1631 AGGATCGCCATTGCCCGGCCCTGGTCCCAACCCCAAGATCCTTTGTGGTAGGGCC 1690
Qy 560 ThrSerAlaLeuAspThrGluSerGluAlaValAlaGlnValAlaLeuAspLysAlaArg 579
Db 1691 ACGTACGCTTGGACACAGAAAGGAAGCCGTGGTTCAGGCCGCTCTGGATAAGGCTAGA 1750
Qy 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaaspVal 599
Db 1751 GAAGGCGGACCAACCATTTGTATAGCTACCGCTTGTCTACAGTGGCAATCTGACGCTC 1810
Qy 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 1811 ATTGCTGGTGTGATGGTGTGCATTGTGGAGCAAGGAATCATCGAAGAGCTCATGAAA 1870
Qy 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 1871 GAGAAGGGCATTTACTTCAAACTTGTCAACACAGACTAGAGGAATGAAATGAAACCA 1930
Qy 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysasp 659
Db 1931 GGAAATTAATGCTTATGAATCCCAAGATGACACTGGTGGCTCTGAGTTGACTTCAGAAAA 1990
Qy 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 1991 TCAAAATCCCTTATA---AGGAGATCAATTCGCGAAGATATCCACAGAGACAAGAC 2047
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
Db 2048 CAGGAGAAGACTTAGTTCCGAAGAGGATGTGGATGAAGATGTGCCTATGCTTCCCTTT 2107
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2108 TGGCAGATCCTAAAGCTAATATTAGTGAATGGCCCTATTTAGTTGGGGTGTACTTTGT 2167
Qy 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2168 GCTGTTATAAATGGGTGCATACACACCGCTTTGGCCATAGTGTTCAAAGATTGTAGGG 2227
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2228 GTTTTTTCAAGACGACGACCATGAAACCAACAAACGGAATGTAACTTGTTCCTCCTT 2287
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2288 CTCTTTCTGGTCAATGGAATGATTTCTTTGTAGTACTTCTTCAAGGCTTCACATTT 2347
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2348 GGCAAGCTGGAGAGATCCTCACCAGCGACTCCGATACATGGTCTTCAAAATCCATGCTG 2407
Qy 800 ArgGluAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2408 CGACAGGATATAAGCTGGTGTGATGACCATGAAACACCAACACCTGGCTGCTGACTACCA 2467
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
|||||

Db 2468 CTCGCTAGTCAGCCTTCTAATGTTAAAGGGGCTATGGGCTCCAGGCTTGTGTAGTTACC 2527
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2528 CAGAAATGTAGCAAAACCTTGGCACAGGAATTATCTTATCTTATGCTATGCTGSCAGCTT 2587
Qy 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2588 ACACCTTTACTTGTAGTAATATATACCACTCATCTCTTGGGTGGAATATTTGAATGAAA 2647
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysLeuGluGlyAlaGlyLysIleAla 899
Db 2648 CTGTGTCTGTGCTCAAGCCTTGAAGACAAGAGCTAGAGATCTCTGGGAAGATCGCT 2707
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2708 ACAGAGCAATTTGAAACTTCCGCACCTGTTCTCTTTGACTAGGAGACAGATTGAA 2767
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2768 ACTATGTATGCCAGAGCTTGCAGATACCATACAGAAATGCTTTGAAAGAACACACGTC 2827
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2828 TTTGGATTCACCTTCGCTTCACCCAGCCATGATTTATTTTCTATGCTGTGTTTC 2887
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 2888 CGGTTCGGTGCTACTTGGTGGCAGAGACTCATGACGTTTGAATAATGTTATGTTGGTA 2947
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 2948 TTTTCTGCTGTGCTTGTGGTCCATGGCAGAGGAATACCACTTCATTCGCTCCCTGAC 3007
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3008 TACCGGAAGGCCAAAGTCTCAGCATCCCATCATCATCAGGATCATTGAGAAAAACCCGAG 3067
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3068 ATTGACAGCTACAGCAGGAGGCTTGAAGCTAATTTGGTTAGAAGAAATGTGAATTT 3127
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3128 AATGAGTTCATGTTCAACTATCCACCCGACCCCAACATCCACAGTCTTCAGGGACTCAGC 3187
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3188 TTCGAGGTGAAGAGGGGCAACCGCTTCGCTGGTGGCAGCAGTGGCTGGGGAAGAGT 3247
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3248 ACAGTGGTCCAGCTGCTCGAGCGCTTACAAACCCCATCGCTGGTGGACAGTGTCTAGAT 3307
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3308 GGCAAGAAATTAACAACTACACGTCCAGTGGCTCCGCGCCACCTGGGCGATTGTCTCC 3367
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3368 CAGAGCCCATCTCTTGTGCTGCGATCACCGAAGACATCGCTACGGAGACAACAGC 3427
Qy 1140 ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisPhe 1159
Db 3428 CGTGTGCTCATGAGGATCGTGGGCGCCGAGGAGGCGCAACATCCACAGCTTC 3487
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3488 ATCAGCTACCTGCTGAGAAATACACACAGAGTGGAGACAAAGGACTCAGCTGCG 3547
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3548 GCGGGCAGAGACGCGCATCGCCATCGCGCGCCCTCGTCGACAGACGCTCACATCTTA 3607

QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
|||||
Db 3608 CTTCTGGATGAAGCGACATCAGCTCTGGATACGGAGAGTGAAGAGTGGTCCAGGAAGCG 3667
QY 1220 LeuAspLysAlaArgGluArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 3668 CTGGACAAAGCCAGGGAAGCGCCACCTGCATTGTGATCGCGCACCGCTGTCCACCATC 3727
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||
Db 3728 CAGAACCGACAGCTTGTGCTGGTGTATTCAGAACCGCCAGGTCAAGAGGACCGCACCCAC 3787
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleThrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 3788 CAGCAGCTGTGCGCCAGAAAGGCATCTATTCTCGATG-----GTTTCAGGCTGGAGCA 3841
QY 1280 LysArg 1281
|||||
Db 3842 AAGCGC 3847
RESULT 11
US-08-461-823-1
; Sequence 1, Application US/08461823
; Patent No. 5593840
; GENERAL INFORMATION:
; APPLICANT: Bhatnagar, Satish K.
; APPLICANT: George Jr., Albert L.
; APPLICANT: Nazarenko, Irina
; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OncorPharm, Inc.
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,823
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/168,621
; FILING DATE: 16-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,433
; FILING DATE: 27-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Karta, Glenn E.
; REGISTRATION NUMBER: 30,649
; REFERENCE/DOCKET NUMBER: PA-0012 CIP 2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301 527-2058
; TELEFAX: 301 208-6997
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-461-823-1
Alignment Scores:
Pred. No.: 0 Length: 2726

Score: 3628.00 Matches: 716
Percent Similarity: 96.28% Conservative: 35
Best Local Similarity: 91.79% Mismatches: 29
Query Match: 56.04% Indels: 0
DB: 1 Gaps: 0
US-09-672-725c-23 (1-1281) x US-08-461-823-1 (1-2726)
QY 502 IleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsnLys 521
|||||
Db 2 ATTGAAAGAGCTCTCAAGGAAGCCAATGCTATGACTTTATCATGAACTGCTCAATAA 61
QY 522 PheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArgIle 541
|||||
Db 62 TTGACACCTCGTTGGAGAGAGAGGGCCAGTTGAGTGGTGGCAGAGACAGAGGATC 121
QY 542 AlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThrSer 561
|||||
Db 122 GCCATTGCACGTGCCCTGGTTGCAACCCCAAGATCCTCTGCTGGATGAGGCCAGTCA 181
QY 562 AlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLysGly 581
|||||
Db 182 GCCTTGGACACAGAAAGCGAGAGTGGTTCAGTGGCTCTGGATAAGCCAGAAAGT 241
QY 582 ArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla 601
|||||
Db 242 CGGACCAACCATGTGATAGCTCATCGTTGCTACAGTTGTAATGCTCAGCTCATCGT 301
QY 602 GlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGluLys 621
|||||
Db 302 GGTTCGATGATGGAGTCATGTTGGAGAAAGGAAATCATGATCACTATGAAGAGAA 361
QY 622 GlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGluAsn 641
|||||
Db 362 GCATTTACTTCAAACTTGTCAATGCGACAGACGAGAAATCAAGTTGAATAGAAAT 421
QY 642 AlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSerGly 661
|||||
Db 422 GCAGCTGATGAATCCAAAGTGAATGATGCTTGGAAATGCTTCAATGATCAAGA 481
QY 662 SerSerLeuLysArgArgSerThrArgSerIleHisAlaProGlnGlyGlnAsp 681
|||||
Db 482 TCCAGTCTAATAAGAAAGATCACTCGTAGAGTGTCCGTGGATCAACGCCCAAGAC 541
QY 682 ArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPheThrPArg 701
|||||
Db 542 AGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCTTTGGAGG 601
QY 702 IleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAlaIle 721
|||||
Db 602 ATTATGAAGCTAAATTAAGTGAATGCTTATTTGTTGGTGTATTTTGTGCCATT 661
QY 722 IleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePhe 741
|||||
Db 662 ATAAATGGAGGCTGCAACGAGCATTTGCAATAATATTTTCAAGATTATAGGGGTTTT 721
QY 742 ThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe 761
|||||
Db 722 ACAAGAAATGATGATCTCTGAAACAAACACAGCAAGATAGTAAGTTGTTTTCATCTGTT 781
QY 762 LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys 781
|||||
Db 782 CTAGCCCTTGGAAATTAATTTCTTTTATTTACATTTTCTTCAGGGTTTCACATTTGGCAAA 841
QY 782 AlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGln 801
|||||
Db 842 GCTGAGAGATCTCTCACAAGGGCTCCGATACATGTTTTCGATCCATGCTCAGACAG 901
QY 802 AspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla 821
|||||
Db 902 GATGTGAGTGGTTGATGACCCCTAAAAACACACACACACACACACACACACACACAC 961
QY 822 AsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn 841

QY	346	SerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaAarGlyAlaAlaTyr	365
Db	1165	AGTTGGGCAACGTCGGCCCAATGGTCAGCGCTTTACGAATGGTGTGCTGCGCGCGG	1224
QY	366	GlutIlePheLleIleAspAsnLysProSerIleAspSerTyrSerLysSerGlyHis	385
Db	1225	AAAGATTACACGAGTTACCGCAGATCGCCACTGGACCCCTATTCTGACGAAGGAAG	1284
QY	386	LysProAspAsnLleLysGlyAsnLeuGluPheLysAsnValHisPheSerTyrProSer	405
Db	1285	GTACTCGACCATTTTGAAGAAATATCGAATTTCGAATGTCAACACATCTACCTTCA	1344
QY	406	ArgLysGluValLysLleLeuLysGlyLeuAsnLysValGlnSerGlyGlnThrVal	425
Db	1345	AGACCGAAGTTACAGTCATGGAAGATGCTCTTTATCGATCCGCGCGGAAGACTACC	1404
QY	426	AlaLeuValGlnAsnSerGlyCysGlyLysSerThrThrValGlnLeuMetGlnArgLeu	445
Db	1405	GCATTGGTGGGCCCATCTGCTCTGGAAGAGTACTGTGTGCGGTGGTGGACGCGTTT	1464
QY	446	TyrAspProThrAspGlyMetValCysLleAspGlyGlnAspIleArgThrIleAsnVal	465
Db	1465	TACCTTCAGTAGGAGCCAGGTATTGCTGGAGCGCATGATATCCAAACCTCAACCTC	1524
QY	466	ArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAlaThr	485
Db	1525	CGTTGGCTGGACAGAGATCTCTCTTGTGACCGAGAACCTGTCTTTTCAGCACCAAG	1584
QY	486	IleAlaGluAsnIleArgTyrGlyArg-----GluAsnValThrMetasp	500
Db	1585	ATCTTTAGAACAATCGAACATGCTTGTATGTGGCACCAAAATTGAGCATGAGTCGAAG	1644
QY	501	GluIle-----GluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMet	516
Db	1645	AAGATTAGAGAGCTGCTTGAGAAATCGGCCAAGTGGCAATGCTCATGATTTATTATG	1704
QY	517	LysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGly	536
Db	1705	GCTCTGCTGAAGGTTACGATACGAATGTGGTCAGCGTGTCTTACTTTTCAGGAGGT	1764
QY	537	GlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeu	556
Db	1765	CAGAAGCAACGATATGCCATGCTGTCGCATTTGTCAGTGACGCCCAAGATTCTGTGCT	1824
QY	557	AspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAsp	576
Db	1825	GATGAGCTACATCAGCTTTGGATACCAAGTCCGAGGCGGTGCTCCAAAGCGCTTTGAT	1884
QY	577	LysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsn	596
Db	1885	AAAGCTCCGAGGTAGAACTACCATTTGTCATTGTCACCGCTTTGTCAACAATCAAAACA	1944
QY	597	AlaaspValIleAlaGlyPheaspaspGlyValIleValGluLysGlyAsnHisAspGlu	616
Db	1945	GCCACAAACATTGTCCTGTCGCGGAAAGATTGCCAACAGGGAACACATGACGAA	2004
QY	617	LeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGlu	636
Db	2005	TTGTGTCATCCGAAGGACGACTACTATAAATCTGTGGAGCGCAACGTATCAACGAGGAG	2064
QY	637	IleGluLeuGluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSer	656
Db	2065	AAGGAAGCAGAA-----GCTCTGGAAGCCGACGCC--GACATGGAC	2103
QY	657	ProLysAspSerGlySerSerLeuIleLysArg-----ArgSerThrArgArg	672
Db	2104	GCCGATGATTTTGGTCAAGAAGGGGTACTGTCATCAAGCTGAGCTTAGCAGCTCGAAT	2163
QY	673	SerIleHisAlaProGlnGlyGlnAspArg-----LysLeuGlyThrLysGlu	688
Db	2164	TCTCTCGATGCTTAGATGAGAAGCGCGCTGGAGATGAAGCAACCGGACGCAAAA	2223
QY	689	Asp-----LeuAsnGluAsnValPro-----ProValSerPhe	699

```
Db 2224 TCGGTTTCAGTCCGCTTCTATTAAGAAAGTCCCGAGCAGTTCGAGAAGTATTCACCTC 2283
      |||:|||||
Qy 700 TrpArgIleLeuLys-----LeuAsnSerThrGluTrpProTyrPheValVal 715
      |||:|||||
Db 2284 TGGACCTCGTCAAGTTCATCGTGCATTAACCCCGCTCGAGTGGTTACATGCTCAAT 2343
      |||:|||||
Qy 716 GlyIlePheCysAlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSer 735
      |||:|||||
Db 2344 GGGTTGACCTTTCGTTCTCGCTTCGCCGGTGGTGTCAACCTACCCAGGCTTTCGTGACGCC 2403
      |||:|||||
Qy 736 ArgIleIleGlyPheThrArgAspGluAspProGluThrLys---ArgGlnAsnSer 754
      |||:|||||
Db 2404 AAGGCAATTAGCAGCTCTCGTTACCCGAATCAATGTTTCACAAACTCAGGCATGACGCG 2463
      |||:|||||
Qy 755 AsnMetPheSerValLeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeu 774
      |||:|||||
Db 2464 AATTCTGCTGCTGTTGCTTCGCTGGTGGAAATGCTCAATTTATCAGCTGTCTATC 2523
      |||:|||||
Qy 775 GlnGlyPheThrPheGlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetVal 794
      |||:|||||
Db 2524 AATGGTACAGCATTTGCTATTGTTCCGAGAGACTCATTCGCCGAGCTAGAAAGTCAAGCG 2583
      |||:|||||
Qy 795 PheArgSerMetLeuArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGly 814
      |||:|||||
Db 2584 TTTAGATCGATTCTTCGTGAGACATCTCATTTTCGACAGGGAAGAGAACACGACCGGT 2643
      |||:|||||
Qy 815 AlaLeuThrThrArgLeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArg 834
      |||:|||||
Db 2644 GCCTTCAGCTCTTCCTTCAACAGACGACGAAGAATCTCTCGGGCGGTTAGCGGAGTGACT 2703
      |||:|||||
Qy 835 LeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIle 854
      |||:|||||
Db 2704 CTGGCACAATCATCATGACTAGCATTGCTGGCGCGCGATGATCATTCATTTGGCG 2763
      |||:|||||
Qy 855 TyrGlyTrpGlnLeuThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGly 874
      |||:|||||
Db 2764 ATCGGATGGAAGCTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2823
      |||:|||||
Qy 875 ValValGluMetLysMetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGly 894
      |||:|||||
Db 2824 TTCTCAGATTCTACATGCTTGTCTCAATTCAGCAACGATCGAAGTCTCCCTACGAGGG 2883
      |||:|||||
Qy 895 AlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArg 914
      |||:|||||
Db 2884 TCTCGGAGCTATGCTTCGGAAGCCACGTCAGCGATCCGCACTGTAGCATCATCTCCTCG 2943
      |||:|||||
Qy 915 GluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeu 934
      |||:|||||
Db 2944 GAACAAGATGCTCGGGCGGTTTACCACGACCGACTACAAAAACAGGCGACGGAAGATTG 3003
      |||:|||||
Qy 935 ArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSer 954
      |||:|||||
Db 3004 ATCTCAGTGTGAGATCCCTCCCTGCTGTATGCTCTCGCGAGCATTTGGTGTCTTCTCG 3063
      |||:|||||
Qy 955 TyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGln 974
      |||:|||||
Db 3064 GTCCGTTGGGCTTCTCGTATGTTGGTGTACACTTCTAGGCCATCATCAGTACAGCATCTTC 3123
      |||:|||||
Qy 975 AspValLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSer 994
      |||:|||||
Db 3124 CGTTCTTCGCTGCTTCTGAGATTCTTTTGGTGGCAATCAGCCCGGAATGTCCTTC 3183
      |||:|||||
Qy 995 SerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIle 1014
      |||:|||||
Db 3184 TCTTTTCCCGGACATAGGCTAGGTAAGAAAGATGCCGCTGCTCAATTCAGAAACATCTTC 3243
      |||:|||||
Qy 1015 GluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGlu 1034
      |||:|||||
Db 3244 GACAGCAAGCAACCAATTTGACATCTGCTCGGATGAGGGCGGAGAAGTTGGAGTCTATGGA 3303
      |||:|||||
Qy 1035 GlyAsnValThrPheAsnGluValPheAsnTyrProThrArgProAspIleProVal 1054
      |||:|||||
```

```
Db 3304 GGCAGAAATCGAATTCGCGGAGCTCCACTTTAGGTACCAACAGCGCGGAGCAGCCCTGTT 3363
      |||:|||||
Qy 1055 LeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSer 1074
      |||:|||||
Db 3364 CTTGAGGAGCTGATTTGAGCGTGAACCTGGACAATACATATGCTTGTGGACCCAGT 3423
      |||:|||||
Qy 1075 GlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGly 1094
      |||:|||||
Db 3424 GGATCGGTAAGACACTACGATTGCTGCTTGGAGCATTTTATGACGCACTGCTGCGA 3483
      |||:|||||
Qy 1095 SerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHis 1114
      |||:|||||
Db 3484 GGGGTCTGCTGTTGACGGAAGGACATACCAAACTCAATGTCAACTATACCGCAGTTTC 3543
      |||:|||||
Qy 1115 LeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAla 1134
      |||:|||||
Db 3544 CTCTCCCTTGTGACCAAGAACCTACTCTGTATCAGGCTACCATCAAGAAAATATCCTG 3603
      |||:|||||
Qy 1135 TyrGlyAspAsnSerArgValValSerHisGluIleMetGlnAlaAlaLysGluAla 1154
      |||:|||||
Db 3604 CTTGAGCTCGATAAGGATGAGCTTTTCGAGGAGACTTTGATTAAAGTCTCAAGATGCC 3663
      |||:|||||
Qy 1155 AsnIleHisHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLys 1174
      |||:|||||
Db 3664 AACATCTATGATTTGCTGTATGCTCACTCCCTGAGGGATTTGACACCGTCTGTGGCAGCAAG 3723
      |||:|||||
Qy 1175 GlyThrGlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArg 1194
      |||:|||||
Db 3724 GGAGCATGTTGCTGTTGGTGACAAAAACAGCGTCTGCCATTGCTCGTCCCTCTCTCGT 3783
      |||:|||||
Qy 1195 GlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLys 1214
      |||:|||||
Db 3784 GACCCCAAGTCTCTTCTGATGAAGCCACATCTGCTTGTGATCCGATCTCGAATCTGAGAA 3843
      |||:|||||
Qy 1215 ValValGlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHis 1234
      |||:|||||
Db 3844 GTGCTGACAGCTGCGCTGGATGCTGCTCCCGCGGGGACACAGATGCTGTGTGCCCCAC 3903
      |||:|||||
Qy 1235 ArgLeuSerThrIleGlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLys 1254
      |||:|||||
Db 3904 CGGCTGAGCACCATTCAAAACGCTGATATAATTTACGTGTTGACCAAGGCAAGATCGTC 3963
      |||:|||||
Qy 1255 GluHisGlyThrHisGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSer 1274
      |||:|||||
Db 3964 GAAAGTGAAGCGCACCGAGTGTGATTGCAAAACAGGCGCGTATTACGAGCTGTGCAAT 4023
      |||:|||||
Qy 1275 ValGlnAlaGlyAlaLys 1280
      |||:|||||
Db 4024 CTCGAGATCTCGGAAG 4041
      |||:|||||
```

RESULT 13

```
US-08-996-545-1
; Sequence 1, Application US/08996545
; Patent No. 5928898
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene at/d of
; TITLE OF INVENTION: Aspergillus nidulans
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/996,545
 FILING DATE:
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
 NAME: Webster, Thomas D.
 REGISTRATION NUMBER: 39,872
 REFERENCE/DOCKET NUMBER: X-11766
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-276-3334

TELEFAX: 317-276-2763

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4002 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..4002

US-08-996-545-1

Alignment Scores:

Pred. No.: 5 31e-274 Length: 4002
 Score: 2482.00 Matches: 536
 Percent Similarity: 59.41% Conservative: 223
 Best Local Similarity: 42.00% Mismatches: 458
 Query Match: 38.34% Indels: 62
 DB: Gaps: 8

US-09-672-725c-23 (1-1281) x US-08-996-545-1 (1-4002)

QY 24 LysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMetPheArgTyrSer 43
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 268 AAGACGAGCTGGAGGAGATCAAGTAACATCTCTTCGTCGTCTCTGGCGGTATGCA 327
 QY 44 AsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleHisGlyAla 63
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 328 ACAAGATGGATATACATTCATGGTAATCAGTACAATCTGTGCCATTCGTGCGGCTGC 387
 QY 64 AlaLeuProLeuMetLeuValPheGlyAsnMetThrAspSerPheAlaAsnAlaGly 83
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 388 ACITTCAGAGGATAATGTTATAT-----411
 QY 84 IleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsnThrGln 103
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 412 -----CAA 414
 QY 104 HisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrSerGlyIle 123
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 415 ATCTCGPACGAGGAGTCTATGATGAATGACCAAGAACGACTACTCTGTTATACCTC 474
 QY 124 GlyAlaGlyValLeuAlaAlaTyrIleGlnValSerPheTrpCysLeuAlaAlaGly 143
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 475 GGTATCGCGGAGTGTGTCATCTGTATGTAGTACTGTGTGGCTTCATCTACCGGAGAA 534
 QY 144 ArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGluIleGly 163
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 535 CAGGCCACGAGAAGATCCGCGAGTATACCTTGAGTCTATCTGTCGCGCAGAACATTGGC 594
 QY 164 TrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspValSerLys 183
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 595 TATTTTGATAAATCGCGTCCGCGGAGTGACCCCGTATACACGCCGATACAAACCTT 654
 QY 184 IleAsnGluIleGlyAspLysIleGlyMetPhePheGlnSerIleAlaThrPhePhe 203
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 655 ATCCAGATGCGAATTCGGAAGAGGTGCTCATTCTGCTTACTCTTGGTACCTTGGGACATTCGG 714
 QY 204 ThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIleLeuAlaIle 223

Db 715 ACAGCATTCATTATCGCCTACGTCAAATACTGGAAGTTGGCTCTAATTTGCAGCTCARCA 774
 ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 224 SerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPheThrAsp 243
 ||| ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 775 ATTGTGGCCCTCGTCTCACCACGCGGTGGTCTCTCAGTTTATCATCAAGTACAGCAAA 834
 QY 244 LysGluLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeuAlaAlaIle 263
 ||| ||| : : : ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 835 AGTCGCTTGACAGCTACGCGTGCAGCGGCACTGTTGCGGAAGAGGTCTACAGCTCCATC 894
 QY 264 ArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyrAsnLysAsnLeu 283
 ||| ||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 895 AGAAATGCCACGCGTTTGSCACCCCAAGACAGCTTCGCAAGCAGTATGAGGTCCACIT 954
 QY 284 GluGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIleSerIleGlyAla 303
 : : : ||||| : : : ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 955 GACGAAGCTGAGAAATGGGAACAAAGAACACAGATTGTATGGGTTTCATGATGGCGCC 1014
 QY 304 AlaPheLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSerLeuVal 323
 ||| ||||| : : : ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1015 ATGTTGGCCTTATGTAATCTGAACTACGCTCTGGCTCTGGATGGGTTCTCGTTTCCTG 1074
 QY 324 LeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerValLeuIleGly 343
 : : : ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1075 GTAGATGGTGCAGTCGATGGGTGATATCTCAGATTCTCATGCCATCTTGATCGGA 1134
 QY 344 AlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArgGlyAla 363
 : : : ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1135 TCGTTCTCTCTGGGAACCTTAGTCCAAATGCTCAAGCATTTTACAAACGCTGTGCCCGC 1194
 QY 364 AlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSerLysSer 383
 ||| : : : ||||| ||||| ||||| : : : ||||| ||||| ||||| ||||| ||||| |||||
 Db 1195 GCCGCAAGATATTTGGAACGATCGATCCCGATCCCATTTAGATCCCATTTTCGAACGA 1254
 QY 384 GlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPheSerTyr 403
 ||| : : : ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1255 GGAAGACGCTCGACCATTTTCAGGCGCCACATTTAGTACGCAATGTCAAGCATATTTAC 1314
 QY 404 ProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSerGlyGln 423
 ||||| ||||| ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1315 CCATCTAGACCGGAGGTACACCGTCATGGAGGATGTTCTCTGTCATCAATGCCCTGGA 1374
 QY 424 ThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeuMetGln 443
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1375 ACAACGCTTTAGTGGCCCTCTGGCTCTGGAAAAGTACGGTGGTGGCTGGTTGGTGGAG 1434
 QY 444 ArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArgThrIle 463
 ||| ||| ||| ||| : : : ||||| : : : ||||| ||||| ||||| ||||| ||||| |||||
 Db 1435 CGATTCTACATGCCCTGTTCGCGGTACGGTTTGTGGTGGATGGCCATGACATCAAGACCTC 1494
 QY 464 AsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAla 483
 ||||| : : : ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1495 AATCTCCGCTGCTCGCCACAGATCTCTTTGGTTAGCCAGGACCTGTTCTTTTGGC 1554
 QY 484 ThrThrIleAlaGluAsnIleArgTyrGly-----ArgGluAsnValThr 498
 ||||| ||| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1555 ACAGCATTTTAAAGAAATATTAGGCACCGGTCTCATCGGCACAAAGTACAGAAATGAATCC 1614
 QY 499 MetAspGlu-----IleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 514
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1615 GAGGATAAGTCCGGAACATCTACGAGAAGCGGCAAAATGCGGAATGCTCATGACTTT 1674
 QY 515 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSer 534
 : : : ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1675 ATTACTGCCCTGCCGAAAGGTTATGAGCAATGTTGGCAGCGGTGCTTCTCTCTTCA 1734
 QY 535 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu 554
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1735 GTGCGCAAGAACGCGCATTCGAATCCCGCTGGCTGTGTAGTGACCAAAATCTCTG 1794
 QY 555 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla 574
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1795 CTCCTGGATGAAGCTACTTCGGCCTTGGACACAAAATCCGAAGCGTGTTCAAGCAGCT 1854
Qy 575 LeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrVal 594
Db 1855 TTGGAGAGGCGACCTGAGAGCGCACTACTATTGTGATCGCTCATCGCTTCCACGATC 1914
Qy 595 ArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValIleGluLysGlyAsnHis 614
Db 1915 AAAACGGCGCACAAATCTGTGTTCTGTCATGTCGCAAAATGTCGAAACAGGAACCTCAC 1974
Qy 615 AspGluLeuMetLysGluLysGlyIleIleThrPheLysLeuValThrMetGlnThrArgGly 634
Db 1975 GATGAATGTGTTGACGCGGAGCGCTTATCGCAAACTGTGTGAGGCTCAACGTATCAAT 2034
Qy 635 AsnGluIleGlu-----LeuGluAsnAlaThrGlyGluSerLysSerGluSerAsp 651
Db 2035 GAACAGAAGAGCTGACGCCCTTGAGGAGCGCGGAGCTGAGGATCTCAGNAATGCAGAT 2094
Qy 652 AlaLeuGluMetSerProLysAspSerGlySerSer-----Leu 664
Db 2095 ATTGCCAAATCAAACTCGCTCAAGCGCATCATCCGATCTCGACGGAAAAACCCACAACC 2154
Qy 665 IleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlnAspArgLysLeu 684
Db 2155 ATTGACCGCAGCGGCACACCAAGCTGTGTTCCAGCGGATTTCTTCAAAGACCC--- 2211
Qy 685 GlyThrLysGluAspLeuAsnGluAsnValProValSerPheThrArgIleLeuLys 704
Db 2212 -----CCCAACAACTCCGAATACTCATTTATGGACGCTGCTCAA 2253
Qy 705 -----LeuAsnSerThrGluThrProThrPheValIleGlyIlePheCysAla 720
Db 2254 TTTGTTGCTTCTTCAACCGCCTGAAATCCGCTACATGCTCATCGCTTCTCTCTCA 2313
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2314 GTGTTAGCTGGTGGTGGCCCAACCCACGCAAGCAGTGTATATGCTAAAGCCATCAGCACA 2373
Qy 741 PheThrArgAspGluAspProGluThrLys---ArgGlnAsnSerAsnMetPheSerVal 759
Db 2374 CTCTCCCTCCCAAGATCAACAATATAGCAAGCTTCGACATGATGCGGATTTCTGGTCAATG 2433
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2434 ATGTTCTGCTGGTGTGATCATTCAGTTTATCAGCGAGTCAACCAATGTGCTGCTCATTT 2493
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2494 GCGGTATGCTCCGAGAGACTTATTCGTCGCGGAGAGACACTGCGCTTCGGACGATATC 2553
Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2554 CGTCAAGACATTGCTTCTTTGACAAGGAAGAGATAGACACCGCGCTCTGACCTCTTTC 2613
Qy 820 LeuAlaAsnAspAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2614 CTGTCCACCGAGACGAGCATCTCTCCGCTGTAGCGGTGTGACTCTAGGCACGATCTTG 2673
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleIleIleIleIle 859
Db 2674 ATGACCTCCACGCCCTAGAGCGGTATCATTTATGCGCTGGGATGGGTGGAAATG 2733
Qy 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2734 GCCTTAGTTGTATCTCGGTGTGCGGTTCTCTCTGGCATGCGGTTCTACCGATCTAT 2793
Qy 880 MetLeuSerGlyClnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2794 ATGCTAGCCCGAGTTTCAATCAGCTCCCAAGCTTGCTTATGAGGGATGCTGCAAACTTGTCT 2853
Qy 900 ThrGluAlaIleGluAsnPheArgThrValSerLeuThrArgGluGlnLysPheGlu 919
Db 2854 TGCAGGCTACATCGTCTATCCGACAGTTGGCTCATTAACCCGGGAAGGATGCTCTGG 2913

Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2914 GAGATTTACCATGCCAGCTTGACGACAAAGGAGGACCACTCTAATCTCTGCTTTGAGG 2973
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2974 TCATPCCCTGTATATGCTGCGCAGGCACTGTGTTCTTCTGCGTTCGCGCTCGGGT 3033
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 3034 TGGTACCGAGGACACTTCTTGTTCACACGAGTAGACATTTTCGCTCTTTGTTGT 3093
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValIleGlnValSerSerPheAlaProAsp 999
Db 3094 TTCTCCGAGATCTCTTGTGTCTCAATCCGCGGACCGCTCTTTCTCTTTCGACACGAC 3153
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3154 ATGGCAAGCGAAGATGCGCGCGCAATTCGAGACTGTTCGACCCGAAAGCCACAA 3213
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3214 ATTGATACCTGCTGAGAGGCGGAGAGCTCGAAACGCTGGAAGTGAATCGAATTT 3273
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3274 AGGAACGTGCACTTCAGATACCGCGCGCCAGACAGCGCTGTCTCGCGCTTGAC 3333
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3334 CTGACCTGGAAGCTTGACAAATATGTGCGCTGTCGGACCCAGCGGTGTGCAAGAGT 3393
Qy 1080 ThrValValGlnLeuLeuArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3394 ACCACATTCATGCTTGAGCGCTTTTACGATCGGATTCGCGGTCATCTTGTGAT 3453
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSer 1119
Db 3454 GGGAGGACATAAGTAACTAAATATCACTCTTACCGCAGCTTCTGTCACTGGTCAGC 3513
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3514 CAGGAGCGGACACTGTACCGGCGCCATCAAGAAACATCTTACTGTTATGTCGAA 3573
Qy 1140 ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3574 GATGACGTACCGGAAGAATTCTTGATTAAAGCTTGCAGGACGCTAATATCTACGACTTC 3633
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3634 ATCATGTGCTCCCGAGGCGTAAATACAGTTGTTGGCAGCAAGGAGGCGATGTTGCT 3693
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3694 GCGCGCCAAAGCAACGTTGGCCATTGCGCGACCTTCTTCGGGATCCCAAAATCCTT 3753
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3754 CTTCGATGAAGCGAGCTCAGCCCTCGACTCCGAGTCAGAAAGGTCGTCGACGCGCT 3813
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3814 TTGATGCGCTGCCGAGGCGCAACCAATCCGCTTGCACACCGACTCAGCAGCAT 3873
Qy 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGlyHisGlyThrHis 1259
Db 3874 CAAAGGCGGACGTTATCTATGTTTTCACCAAGCAAGATCGTCGAAAGGGAACGAC 3933
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3934 AGCGAACTGGTCCAGAAAAAGGCGGTACTACGAGCTGTCTCAACTTTCAGAGCTGGGC 3993

```

Oy 1280 Lys 1280
Db 3994 AAG 3996

RESULT 14
US-08-996-545-3
; Sequence 3, Application US/08996545
; Patent No. 5928898
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene atrd of
; TITLE OF INVENTION: Aspergillus Nidulans
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,545
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11766
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-2763
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-996-545-3

Alignment Scores:
Pred. No.: 5,31e-274 Length: 4002
Score: 2482.00
Percent Similarity: 59.41%
Best Local Similarity: 42.00%
Query Match: 38.34%
DB: 2
Gaps: 8

US-09-672-725C-23 (1-1281) x US-08-996-545-3 (1-4002)
Oy 24 LysAsnGluLysLysGluLysLysProThrValSerThrPheAlaMetPheArgTyrSer 43
Db 268 AAGACCGAGUGGAGGAGAGUAAAGUAAACAUCCUUCUGGUCUGCGCGGUAUGCA 327
Oy 44 AsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleHisGlyAla 63
Db 328 ACAAAGAUGGAUAUACUUAUGAUGUAUAUAGUACAUAUGUGCAUAUGCGCGGUGC 387
Oy 64 AlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAsnAlaGly 83
Db 388 ACUUCACAGAGGAUAUAGUUAUAU-----
Oy 84 IleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsnThrGln 103

```

```

Db 412 -----CAA 414
Oy 104 HisPheIleAsnHisLeuGluMetThrThrTyrAlaTyrTyrSerGlyIle 123
Db 415 AUCUCUACGACGAGUUAUGAUGAACCAAGACAGUACUGUACUUAUCCUACUC 474
Oy 124 GlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeuAlaGly 143
Db 475 GGUAUGCGGAGUUGUCACUGUCUUAUGUAGUACUUGUUGUUCUAUACCGGAGAA 534
Oy 144 ArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGluIleGly 163
Db 535 CAGCGCACGACGAGAACUCCGAGUUAUACCUUAGUUCUCCGCGCAGAACAUUGGC 594
Oy 164 TrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspValSerLys 183
Db 595 UAUUUGUAUAAACUCUGGCGCGGGAAGUACACCCGUUAUAAACAGCCCAACACCU 654
Oy 184 IleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAlaThrPhePhe 203
Db 655 AUCCAGGAUGGCAUUCGAGAGGCGGUCUCACUUAUGACUCCUGCGACAUUCGUG 714
Oy 204 ThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIleLeuAlaIle 223
Db 715 ACAGCAUUCAUUACGCUACGUAACUAGGAGUUGGCUAAUUGCAGCUCACAACA 774
Oy 224 SerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPheThrAsp 243
Db 775 AUUGGGCCUCUCUACCAUGGCGGUGGUCUACAGUUAUACUAAAGACAGCAAA 834
Oy 244 LysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaIleGluValLeuAlaAlaIle 263
Db 835 AAGUCGCUUACAGACGUCGAGCGGCGACUGUUGCGGAGAGAGUACUACAGCUCCAU 894
Oy 264 ArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluAlaTyrAsnLysAsnLeu 283
Db 895 AGAAAGUCCACAGCGUUGGCGACCAAGACAAAGCUUUGCGAGCAGUAGGAGUCCACUA 954
Oy 284 GluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSerIleGlyAla 303
Db 955 GACGAAGCUGAGAAAUUGGGGACAAAGACAGAUUGUUGUUGUUGUUGGCGGC 1014
Oy 304 AlaPheLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSerLeuVal 323
Db 1015 AUGUUGGCUUAUGUACUAGCAGACUACGUCUUGGCUUUGGAGGUCUUGUUGUUG 1074
Oy 324 LeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerValLeuIleGly 343
Db 1075 GUAGAUGGUGCAGUCGAGUGGUGUAUUCUCACAGUUCUACUAGGCCAUUCGUAUCGA 1134
Oy 344 AlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArgGlyAla 363
Db 1135 UCUUCUUCUUGGGGAGCAGUUAUGUACAAUAGCUCUACAGCAUUAUACACCGUGGCCGCG 1194
Oy 364 AlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSerLysSer 383
Db 1195 GCGCAAGAUUAUUGGAGCAGUACCGUCCAGUCCUCCAUUAUAGUACCAUUAUUCGAGCA 1254
Oy 384 GlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPheSerTyr 403
Db 1255 GGGAGACCGUCCACCAUUAUGAGGCGCACAUUGAGUUAUCGCAUUGCAUUAUUAUAC 1314
Oy 404 ProSerArgLysGluValLysIleLeuLysGlyLysLeuAsnLeuLysValGlnSerGlyGln 423
Db 1315 CCAUCUAGACCGAGGUCACCGUACUAGGAGAUUUUCUCUACUACUACUCCGCGCGGAGAAA 1374
Oy 424 ThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeuMetGln 443
Db 1375 ACAACCGCUUAUAGUCGCGCCUUCUGGCUAGGAGAAAGUACGGUGGUGGUGGUGGUGAG 1434
Oy 444 ArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArgThrIle 463

```

Db 1435 CGAUUCUACAUGCCUUGUGCGGUAUGCGGUUUUUUGUGGAGGCGCAUAGCAUACAAGGACUUC 1494
Qy 464 AsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAla 483
Db 1495 AAUUCUCCGUGGUGGCAACAGAUUCUUGGUUAGCCAGGAGCCUUGUUCUUUUUGGC 1554
Qy 484 ThrThrIleAlaGluAsnIleArgTyrGly-----ArgGluAsnValThr 498
Db 1555 ACCGAGAUUAUAGAAUAUAGGACGCGUUCUACCGGCACAAAGUACGAGAAUAGAUCC 1614
Qy 499 MetAspGlu-----IleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 514
Db 1615 GAGGAUAGGUGGCGGGAACUCUACGAGAACGCGCAAAAUUGCGGAUUGCUAGCAUUC 1674
Qy 515 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSer 534
Db 1675 AUUACUGCCUUGCCUAGAGUUAUGAGACCAUUGUUGGCGGUGGUUUCUUCUUA 1734
Qy 535 GlyGlyGlnLysGlnArgIleAlaIleAlaIleAlaLeuValArgAsnProLysIleLeu 554
Db 1735 GUGGCGCAAGAAACAGCAUUGCAUUGCGGCGUUGUUGAGUGAGCCCAAAAUCCUG 1794
Qy 555 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla 574
Db 1795 CUCCUGGAUGAAGCUUUCUGGCCUUGGACACAAAUUCCGAGGCGUGUUAAGCAGCU 1854
Qy 575 LeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrVal 594
Db 1855 UUGGAGGGGACGUGGAGGGCGGAACUACUUAUUGUUGUUGUUGUUGUUGUUGUUGUUG 1914
Qy 595 ArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHis 614
Db 1915 AAAACGGGCGACAAUUGUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUG 1974
Qy 615 AspGluLeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGly 634
Db 1975 GAUGAAUUGGUUGACCGGGGCGCCUUAUCCCAACUUGUGGAGGCUACAGUAUCAA 2034
Qy 635 AsnGluIleGlu-----LeuGluAsnAlaThrGlyGluSerLysSerGluSerAsp 651
Db 2035 GAACAGAAGGAAGUAGCGCCUUGGAGGACCGCGAGGAGGAGGAGGAGGAGGAGGAGG 2094
Qy 652 AlaLeuGluMetSerProLysAspSerGlySerSer-----Leu 664
Db 2095 AUUGCCAAAUAUCAAUUGCGUACGAGCGCAUACUCCGUAUCUGGAGGAGAAACCCACAC 2154
Qy 665 IleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAspArgLysLeu 684
Db 2155 AUUGACCGGACGCGGACCCACAAUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUGUUG 2211
Qy 685 GlyThrLysGluAspLeuAsnGluAsnValProProValSerPheThrArgIleLeuLys 704
Db 2212 -----CCCGAAACAAUCCGAAAUACUUAUUGGAGGAGGAGGAGGAGGAGGAGG 2253
Qy 705 -----LeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db 2254 UUUUGUUCUUCUUAACCGCCUUAUCCGUAUCCGUAUCCGUAUCCGUAUCCGUAUCCGUA 2313
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleClyIle 740
Db 2314 GUGUAGCUGGUGGUGGCGCAACCCAGCAAGCAGGUGCUUAUUGCUUAAAGCCUACGAGC 2373
Qy 741 PheThrArgAspGluAspProGluThrLys---ArgGlnAsnSerAsnMetPheSerVal 759
Db 2374 CUUCGCGUCCAGAAUACAUAUAGCAAGCUUUCAGCAUGAUGGAGUUAUUGGUGUUAUUG 2433
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2434 AUGUUCUUGGUGGUGUUAUACAUGUUAUACAGUUAUACAGUUAUACAGUUAUACAGUUA 2493
Qy 780 GlyLysAlaGlyIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2494 GCCGUAGUUCGCGAGAGACUUAUUGCGGCGGAGAGCAGCAGCAGCAGCAGCAGCAGC 2553

Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2554 CGUACAGACAUUGCUUUUUGACAGGAAGAAUAGCACCAGCGCUCUAGCUCUUCUUC 2613
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2614 CUGUCCACCGAGCAAGCAUCUCCGUGUUGAGGCGUGAGUCACUAGGACGACAUUG 2673
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2674 AUGACUCCAGCAGCCUAGAGCGGCUUAUUAUUGCCUUGGCGGAUUGGUGGAAAUUG 2733
Qy 860 ThrLeuLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLys 879
Db 2734 GCCUUAUUGUUAUCUGGUGUUGCGGCUUCUGGCAUGCGGCUUACCGAUUCUUAU 2793
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2794 AUGUAGCCCGAUUUAUACACGCUUUAUUGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2853
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2854 UGCAGGCUACUUGCUUACCGCAGUUGCGUCAUAUACCCGGAAGGAGGAGGAGGAGG 2913
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2914 GAGAUUUAUACUAGCCAGCUUUGAGCGACAAAGGAGGAGGAGGAGGAGGAGGAGG 2973
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2974 UCAUCCUGUUAUUGGCGUGCGAGGACUUAUUGUUGUUGUUGUUGUUGUUGUUGUUG 3033
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 3034 UGUACGAGGAGGACACUUCUUGGUCACCGAGAGUAGCAUUAUUCGCUUUGUUGUUG 3093
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3094 UUCUCCAGAUUCUUCUUGGUCUUAUCCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 3153
Qy 1000 TyrAlaLysAlaLysValSerAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3154 AUGGCAAGCGGAAGUUGCGCGCGGAGUUCGAGAGUUCGAGAGGAGGAGGAGGAGG 3213
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3214 AUUGAAUACUGGUGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3273
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnLysLeuSer 1059
Db 3274 AGNACGUGGACUUCAGAUACCCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3333
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3334 CUGACCGUGAAGCGUGGACAAUUGUUGCGUUGCGAGGAGGAGGAGGAGGAGGAGGAG 3393
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3394 ACCACCAUGCAUUGUUGAGCGCUUUAUAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3453
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3454 GGAAGGACAAUAGAAUAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3513
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3514 CAGGAGCGGACACUGUAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3573
Qy 1140 ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisPhe 1159
Db 3574 GAUGACGCGGAGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3633

QY 304 AlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheThrTyrGlyThrSerLeuVal 323
 Db 1015 ATGTTGGCCTTATGACGAACTACGGTCTGGCTTGGATGGTCTCTGCTTCTCCTG 1074
 QY 324 LeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerValLeuIleGly 343
 Db 1075 GTAGATGGTCCAGTCGATGGGTGATATCTCACAGTTCTCATGGCCATCTGTATCGGA 1134
 QY 344 AlaPheSerIleGlyGlnAlaSerProSerIleGlyAlaPheAlaAsnAlaArgGlyAla 363
 Db 1135 TCGTTCTCTTGGGAACGTTAGTCCAAATGCTCAAGCATTTTACAAACGCTGTGGCCGCG 1194
 QY 364 AlaTyrGluIlePheIleLeuLeuAlaSerProSerIleAlaSerTyrSerIleSer 383
 Db 1195 GCGCAAAAGATTTTGGAAACGATCGATCCGACGATCCCAATAGATCCCATATTCGAACGAA 1254
 QY 384 GlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPheSerTyr 403
 Db 1255 GGAAGACGCTGACCATTTTGGGGCCACATGTAGTTACGCAATGTCAAGCATATTTAC 1314
 QY 404 ProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSerGlyGln 423
 Db 1315 CCATCTAGACCCGAGTCCAGCTCATGGAGGATGTTCTCTCTCAATGCGCGCTGGA 1374
 QY 424 ThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeuMetGln 443
 Db 1375 ACACCGCTTGTAGTCGGCCCTCTGGCTCTGGAAGATGACGGTGGCTTGGTTGAG 1434
 QY 444 ArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArgThrIle 463
 Db 1435 CGATTCTACATGCTTTCGCGGTACGGTTTTCCTGGATGGCCATGACATCAAGGACCTC 1494
 QY 464 AsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAla 483
 Db 1495 AATCTCGCTGGCTTCGCAACAGATCTCTTGGTTAGCCAGGACCTGTTCTTTTGGC 1554
 QY 484 ThrThrIleAlaGluAsnIleArgTyrGly-----ArgGluAsnValThr 498
 Db 1555 ACACGATTTATGAATATTTAGCAGCGTCTCATCGGCACAAATGACGAGAAATGATCC 1614
 QY 499 MetAspGlu-----IleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 514
 Db 1615 GAGGATAAGGTCGGGAACCTCATCGAGACGCGGCAAAATGGCGAATGCTCATGACTT 1674
 QY 515 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSer 534
 Db 1675 ATTACTGCTTGCCTGAAGTTATGAGACCAATGTTGGCGACGCTGCTCTCTCTTCA 1734
 QY 535 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu 554
 Db 1735 GGTGGCCAGAAACAGCGCATTCGAATCGCCGCTGCTGTTAGTGACCCCAAAATCTCG 1794
 QY 555 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla 574
 Db 1795 CTCTGGATGAAGCTTATCGGCTTGGACACAAATCCGAGGCGTGGTTCAAGCAGCT 1854
 QY 575 LeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrVal 594
 Db 1855 TTGGAGAGGGCAGCTGAAGGCGGAATCTATTGTGATCGCTCATCGCTTTCACGATC 1914
 QY 595 ArgAsnAlaAspValIleAlaGlyPheAspGlyValIleValGluLysGlyAsnHis 614
 Db 1915 AAAACGGCCACAACTATGTTGTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1974
 QY 615 AspGluLeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGly 634
 Db 1975 GATGAATGGTTGACCGGGGCGCTTATCGCAAACTTGTGGAGGCTCAAGCTATCAAT 2034
 QY 635 AsnGluIleGlu-----LeuGluAsnAlaThrGlyGlnLysSerGluSerAsp 651
 Db 2035 GAACAGAAAGGAGCTCAGCGCTTGGAGACCGCAGCGCTGAGGATCTCACGAATCGCAT 2094
 QY 652 AlaLeuGluMetSerProLysAspSerGlySer-----Leu 664

Db 2095 ATTGCCAAATCAAAACATCGCTCAAGCGCATCATCCGATCTCGACGGAACCCACAACC 2154
 QY 665 IleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAspArgLysLeu 684
 Db 2155 ATTGACCGCAGCGGACCCACCAAGTCTGTTCACGCGGATCTTCTTAAAGACCC--- 2211
 QY 685 GlyThrLysGluAspLeuAsnGluAsnValProValSerPheThrArgIleLeuLys 704
 Db 2212 -----CCGGAACAACATCCGAAATATCATTTATGAGCGCTCTCAAA 2253
 QY 705 -----LeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
 Db 2254 TTTGTGCTTCTTCAACCGCCTGAAATCCGATACATGCTCATCGTCTGTCTTCTCA 2313
 QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
 Db 2314 GTGTAGCTGTGTGGCCAAACCCACGACGATGCTATATGCTAAAGCCATCAGCACA 2373
 QY 741 PheThrArgGluAspProGluThrLys---ArgGlnAsnSerAsnMetPheSerVal 759
 Db 2374 CTCTCGCTCCAGAAATCACAATATAGCAAGCTTCGACATGATGCGGATTTCTGGTCATTG 2433
 QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
 Db 2434 ATGTCTCTCGGTGGTGTATCATTTACGTTTATCAGCAGTCAACCAATGTGTGCTGCAATT 2493
 QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
 Db 2494 GCGGTATGCTCCGAGACATTTATTCGTCGCGGAGAGACACTGCCITTCGAGGACTACTC 2553
 QY 800 ArgGlnAspValSerThrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
 Db 2554 CGTCAAGACATTTGCTTTTACAAAGAAAGAGAATAAGCACCGCGCTCTGACTCTCTTC 2613
 QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
 Db 2614 CTGTCCACGAGAGACGACATCTCTCGGTGTAGCGTGTGACTCTAGGACGATCTTG 2673
 QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTyrGlnLeu 859
 Db 2674 ATGACCTCCAGACCTTAGGCGGCTATCATTTATGCTCCGCTGGGATGGTGGAATG 2733
 QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
 Db 2734 GCCTTAGTTGTATCTCGTGTGCGGTCTCTCTGTCATCGCTTCTACCGATCTAT 2793
 QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluAlaGlyLysIleAla 899
 Db 2794 ATGTAGCCCGATTTCAATCAGCTCCAAAGCTTGTCTTATGAGGATCTGCATACTTTGCT 2853
 QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
 Db 2854 TGCAGGCTACATCGTCTATCCGACAGTTCGCTCATTAACCCGGNAAGGATGTCTGG 2913
 QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
 Db 2914 GAGATTTACCATGCCAGCTTGACGCAACAGGAGGACCAAGTCTAATCTCTGTCTGAGG 2973
 QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
 Db 2974 TCATCCCTGTTATGCTGCGAGGACCTGTTTCTCTGCTGCTGCTGCTGCTGCTGCTG 3033
 QY 960 ArgPheGlyAlaTyrLeuValAlaAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
 Db 3034 TGGTACGGAGGACACTTCTTGTGTACCGAGTATGACATTTTCGCTCTTGTGTGTGT 3093
 QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
 Db 3094 TTCTCCGAGATCTCTTGTGTGCTCAATCGCGGCGACCGCTCTTTTCTTTCACCCAGAC 3153
 QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019

```

Db 3154 ATGGCAAGGCGAAGATCGCGCGCCGAATTCGACGACTGTTCGACCGAAGCCACAA 3213
Qy 1020 ileaspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3214 ATTGATAACTGGTCTGAAGAGGGCGAGAGCTCGAAACGGTGGAGGTGAAATCGAATTT 3273
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3274 AGGAACGTGCACCTTCAGATACCCGACCGCCAGACAGACCTGTCTCGCGGGCTTGGAC 3333
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3334 CTGACCGTGAAGCCCTGGACAATATGTTGCGCTTCTCGGACCCAGCGTTGTGCAAGAGT 3393
Qy 1080 ThrValValGlnLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleasp 1099
Db 3394 ACCACCATGTGATCTTGAGCGCTTTTACGATGCGGATGCGGGTCCATCCTTGTGTGAT 3453
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3454 GGGAGGACATAAGTAACTAATATCAACTCTACCGCAGCTTCTGTCACTGGTCAGC 3513
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3514 CAGGAGCGACACTGTACAGGCGACCATCAAGGAAACATCTTACTTGGTATTGTCGAA 3573
Qy 1140 ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3574 GATGACGTACCGGAGGAATCTTGATTAAGGCTTGCAAGGACGCTAATATCTACGACTTC 3633
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3634 ATCATGTGCTCCCGAGGGCTTTAATACATGTTCTTGGCAGCAGGAGGCATGTTGTCT 3693
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3694 GCGGCCCAAAAGCAACGTGTGCCATTGCGCGAGCCCTTCTTCGGGATCCCAAAATCCTT 3753
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3754 CTTCTCGATGAGCGACGTACGCCCTCGACTCCGAGTCAGAAAAGGTGTCGACGGCGCT 3813
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3814 TTGGATGCGCTGCCCGAGGCCGAACCAACATCCCGCTTGCACACCGACTCAGCAGATT 3873
Qy 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3874 CAAAAGCGGACGTTATCTATGTTTTCGACCAAGCAAGATCGTCGAAAGCGGAACGCAC 3933
Qy 1260 GlnGlnLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3934 AGCGAATGGTCCAGAAAAGGGCGGTACTACGAGCTGGTCACTTGCAGAGCTTGGGC 3993
Qy 1280 Lys 1280
Db 3994 AAG 3996

```

Search completed: November 6, 2002, 19:23:29
Job time : 435.556 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 6, 2002, 18:33:09 ; Search time 20.063 Seconds
(without alignments)
6135.192 Million cell updates/sec
Title: US-09-672-725C-23
Perfect score: 6474
Sequence: 1 MDPEGGRKGSAEKNFKMGK.....LLAQKGIYFSMYSVQAGAKR 1281
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*

- 1: pir1.*
- 2: pir2.*
- 3: pir3.*
- 4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5866	90.6	1280	1 DVHU1	multidrug resistan
2	5656.5	87.4	1276	1 DVHY1C	multidrug resistan
3	5638	87.1	1276	2 A34786	multidrug resistan
4	5301	81.9	1276	1 DVMS1	multidrug resistan
5	5231.5	80.8	1277	2 JH0502	p-glycoprotein - r
6	4995	77.2	1104	1 DVMS1A	multidrug resistan
7	4910.5	75.8	1279	1 DVHU3	multidrug resistan
8	4861	75.1	1276	1 DVMS2	multidrug resistan
9	4823.5	74.5	1281	2 I48123	multidrug resistan
10	4808	74.3	1278	2 S41646	p-glycoprotein iso
11	4427	68.4	1287	2 S55692	p-glycoprotein - r
12	3214.5	49.7	1321	2 T42228	multidrug resistan
13	3210.5	49.6	1321	2 T42228	p-glycoprotein sis
14	2862	44.2	1294	2 T19982	bile salt transpor
15	2824	43.6	1289	2 D87789	hypothetical prote
16	2823.5	43.6	1275	2 T31073	protein C34G6.4 [i
17	2746	42.4	1321	2 T23476	multidrug resistan
18	2744	42.4	1321	2 S27337	hypothetical prote
19	2674.5	41.3	655	1 DVHY2C	multidrug resistan
20	2643	40.8	1283	2 A47377	multidrug resistan
21	2607.5	40.3	1286	2 T02187	probable ABC trans
22	2601	40.2	1292	2 T48007	probable ABC trans
23	2594.5	40.1	1282	2 E86155	probable ABC trans
24	2562.5	39.6	1302	2 A41249	multidrug resistan
25	2526	39.0	1229	2 D85023	p-glycoprotein-lik
26	2523	39.0	1229	2 T52319	p-glycoprotein-lik
27	2523	39.0	1230	2 E85023	probable p-glycopr
28	2493.5	38.5	1302	2 B41249	multidrug resistan
29	2478	38.3	1229	2 F86155	probable ABC trans

30	2446	37.8	1408	2 T43261	multidrug resistan
31	2442.5	37.7	1323	2 H85202	hypothetical prote
32	2428	37.5	1310	2 S30328	multidrug resistan
33	2423	37.4	1286	2 A42150	p-glycoprotein ppp
34	2407	37.2	1266	2 T22090	hypothetical prote
35	2398	37.0	1268	2 T22094	hypothetical prote
36	2385	36.8	1302	2 S30327	multidrug resistan
37	2345	36.2	1254	2 S27338	p-glycoprotein C -
38	2335.5	36.1	1222	2 T14805	hypothetical prote
39	2299	35.5	1233	2 T04251	probable p-glycopr
40	2298.5	35.5	1245	2 G86404	multidrug resistan
41	2294	35.4	1254	2 T30855	multidrug resistan
42	2288	35.3	1307	2 T30882	multidrug resistan
43	2275	35.1	1362	2 T41534	leptomycin B resis
44	2270.5	35.1	1318	2 T21266	hypothetical prote
45	2247	34.7	1247	2 F86405	probable p-glycopr

ALIGNMENTS

RESULT 1

DVHU1
multidrug resistance protein 1 - human
N:Alternate names: p-glycoprotein 1
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1990 #sequence_revision 18-Aug-1995 #text_change 19-Jan-2001
C:Accession: A34914; PS0162; S15500; A25059; S43838; I52238; I65204
R:Chen, C.; Clark, D.; Ueda, K.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.
J. Biol. Chem. 265, 506-514, 1990
A>Title: Genomic organization of the human multidrug resistance (MDR1) gene and origin
A:Reference number: A34914; MUID:90094448
A:Accession: A34914
A:Molecule type: DNA
A:Residues: 1-1280 <CHE>
A:Cross-references: GB:M29447; GB:J05168; NID:g187496; PIDN:AAA59576.1; PID:g386862
R:Kioka, N.; Yanano, Y.; Komano, T.; Ueda, K.
submitted to JIPID, April 1991
A:Reference number: PS0162
A:Accession: PS0162
A:Molecule type: DNA
A:Residues: 1-22 <KIO>
R:Kioka, N.; Yanano, Y.; Komano, T.; Ueda, K.
submitted to the EMBL Data Library, April 1991
A:Description: transcriptional regulation of multidrug resistance gene (MDR1) express

A:Reference number: S15500
A:Accession: S15500
A:Molecule type: DNA
A:Residues: 1-22, 'R' <KI2>
A:Cross-references: EMBL:X58723; NID:g34522; PIDN:CAA41558.1; PID:g34523
R:Chen, C.; Chlin, J.E.; Ueda, K.; Clark, D.P.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.
Cell 47, 381-389, 1986
A>Title: Internal duplication and homology with bacterial transport proteins in the m
A:Reference number: A25059; MUID:87028230
A:Accession: A25059
A:Molecule type: mRNA
A:Residues: 1-184, 'V', 186-1280 <CH2>
A:Cross-references: GB:M14758; NID:g187468; PIDN:AAA59575.1; PID:g307180
R:Chambers, T.C.; Pohl, J.; Glass, D.B.; Kuo, J.F.
Biochem. J. 299, 309-315, 1994
A>Title: Phosphorylation by protein kinase C and cyclic AMP-dependent protein kinase
A:Reference number: S43838; MUID:94220047
A:Accession: S43838
A:Molecule type: protein
A:Residues: 656-689 <CHA>
R:Geckeler, V.; Weger, S.; Probst, H.
Biochem. Biophys. Res. Commun. 169, 796-802, 1990
A>Title: mdrl/p-glycoprotein gene segments analyzed from various human leukemic cell
A:Reference number: I52238; MUID:90290529
A:Accession: I52238
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 178-215 <RES>

R:Devault, A.; Gros, P.
Mol. Cell. Biol. 10, 1652-1663, 1990
A>Title: Two members of the mouse mdr gene family confer multidrug resistance with over
A:Reference number: A34786; MUID:90205845
A:Accession: A34786
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1276 <DEV>
A:Cross-references: GB:M30697; NID:g199111; PIDN:AAA39517.1; PID:g387429
R:Hsu, S.I.H.; Cohen, D.; Kirschner, L.S.; Lothstein, L.; Hartstein, M.; Horwitz, S.B.
Mol. Cell. Biol. 10, 3596-3606, 1990
A>Title: Structural analysis of the mouse mdr1a (P-glycoprotein) promoter reveals the ba
A:Reference number: A35671; MUID:90287150
A:Accession: A35671
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-525, 'HV', 528-938, 'S', 940-1035, 'V', 1037-1276 <HSU>
A:Cross-references: GB:M33581; NID:g199104; PIDN:AAA39514.1; PID:g387427
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; membrane protein; nucleotide binding; P-loop
F:406-600/Domain: ATP-binding cassette homology <ABC1>
F:423-431/Region: nucleotide-binding motif A (P-loop)
F:547-551/Region: nucleotide-binding motif B
F:1049-1245/Domain: ATP-binding cassette homology <ABC2>
F:1068-1074/Region: nucleotide-binding motif A (P-loop)
F:1192-1196/Region: nucleotide-binding motif B
F:429/Binding site: ATP (Lys) #status predicted
F:1072/Binding site: ATP (Lys) #status predicted

Query Match 87.1%; Score 5638; DB 2; Length 1276;
Best Local Similarity 87.1%; Pred. No. 0;
Matches 1116; Conservative 74; Mismatches 84; Indels 8; Gaps 4;

QY 1 MDPEGGRKSAEKNFMKMKKKKKEKPKTSTVFAMFRYSNNLDRLMLVGMTAAII 60
DB 1 MELEEDLKRADKNFSKMGKKKK -EKKEKPAVSLVIMFRAGMDRLMLVGMTAAII 59

QY 61 HGAALPLMLLVGNMTDFSPANAG-ISRNKTFPVIINESITNTTQHFIFNHEEMTYAY 119
DB 60 HGVALPLMLLIFGDMTDFSPASGVNSKNST-----NMSEADKRAMFAK-LEBEEMTYAY 113

QY 120 YSIGAGVLVAAYIQVSWFLAAGRQILKIRKQFHAIMRQIGFVDHVGELNRLTD 179
DB 114 YTGIGAGVLVAYIQVSWFLAAGRQILKIRKQFHAIMRQIGFVDHVGELNRLTD 173

QY 180 DVSKINEGIDKIGMFQSIATFFFGFIVGFTGKWLTVILAIISPVLGLSAAIWAKILS 239
DB 174 DVSKINEGIDKIGMFQAMATFFGGFIIGFTGKWLTVILAIISPVLGLSAAIWAKILS 233

QY 240 SFTDKELAYAKAGAAVEVLAAIRTVIAFGQKKELERYKNLBEAKGIGIKKAITANI 299
DB 234 SFTDKELAYAKAGAAVEVLAAIRTVIAFGQKKELERYNNLEAKRLGIKKAITANI 293

QY 300 STGAPELLIYASVALAFWGTSLVLSSEYSIQOVLTVFVSFLIGAFSGQASPTAEAFAN 359
DB 294 SMGAPELLIYASVALAFWGTSLVLSKEYSIQOVLTVFVSFLIGAFSGQASPTAEAFAN 353

QY 360 ARGAAVEIFKIDNKPSIDYSKSGHKPDNIKGNLEFNHPSYPSRKEVKILKGLNLKV 419
DB 354 ARGAAVEIFKIDNKPSIDFSKSGHKPDNIOGNLEFNHPSYPSRKEVQILKGLNLKV 413

QY 420 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDINTINVRHLREITGVVSGEP 479
DB 414 KSGQTVLVGNSGCGKSTTVQLMQRLYDPLDGMVSDIGQDINTINVRVLYREIIGVSGEP 473

QY 480 VLFATTIAENIRYGRNVTMDIEKAVKANAYDFIMKLPNKFDTLVGERGAQLSGGQK 539
DB 474 VLFATTIAENIRYGRNVTMDIEKAVKANAYDFIMKLPNKFDTLVGERGAQLSGGQK 533

QY 540 RTAIALRVNRPKILLDLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 599
DB 534 RTAIALRVNRPKILLDLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 593

QY 600 IAGFDDGVIVVEKGNHDELMKEGIYFKLVLTMTQTRGNEIELENATGESKSESDALEMSPKD 659
DB 594 IAGFDGGVIVVEQGNHDELMREKGIYFKLVMTQTAGNETELGNEACKSKDEIDNLDMSKD 653

QY 660 SGSSLIKRSTRSRSHAPOGQDRKLGTKEDLNENVPVPSFWRILKLNSTEWPFVVGIFC 719
DB 654 SGSSLIKRSTRKSCICGPHDQDRKLTKEALDEDPVPSFWRILKLNSTEWPFVVGIFC 713

QY 720 AIINGGLQPAFSIIFSRIGITRDEDEPKTRONSMFSLVFLVIGTISFTFFPLOGFTF 779
DB 714 AIINGGLQPAFSVFSKVGVTGGPQETORQNSNLSFLFLILGILISITFFPLOGFTF 773

QY 780 GKAGEILTRLRYMVFKSMRLQDYSWFDKPKNTTGTALTRLANDAAQVKAIGASRLAVIT 839
DB 774 GKAGEILTRLRYMVFKSMRLQDYSWFDKPKNTTGTALTRLANDAAQVKAIGASRLAVIT 833

QY 840 QNIANLGTGIIISLIYGWQLTLLLLLAIVPIIAIAGVVMKMLSGOALKDKKELEGAKIA 899
DB 834 QNIANLGTGIIISLIYGWQLTLLLLLAIVPIIAIAGVVMKMLSGOALKDKKELEGAKIA 893

QY 900 TEAENFTVVSITREOKFEYMAQSLQVPYRNSLRKAHIFGVSFSTQAMMVFSAACF 959
DB 894 TEAENFTVVSITREOKFEYMAQSLQVPYRNSLRKAHIFGVSFSTQAMMVFSAACF 953

QY 960 RFGAYLVANPEPMFQDVLVLSAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPL 1019
DB 954 RFGAYLVLTQQLMTPENVLVLSAIVFGAMAVGVSSFPADYAKAKVSAASHIRIETPE 1013

QY 1020 IDSYSPHGLKPNTELEGNTVNEVFNTPRPDIPVLOGLSLEVKKGTALVGVSSGCGKS 1079
DB 1014 IDSYSTGCLKPNMLEGNVSGFVFNTPRPSPVLOGLSLEVKKGTALVGVSSGCGKS 1073

QY 1080 TVQOLLERFYDPLAGSVLIDGKEIKHLNVWLRAHLGIVSQEPIFLDCSTAEINAYGDNS 1139
DB 1074 TVQOLLERFYDPMAGSVFLDGEKEIKQLNVQWLRAQLGIVSQEPIFLDCSTAEINAYGDNS 1133

QY 1140 RVVSHEEIMQAQKANIHFETLPEKYNTRVGDGKQQLSGGQKQRIAIARALVRPHIL 1199
DB 1134 RVVSYEETVRAAKEANIHQFIDSLPKYNTYRVDGKQQLSGGQKQRIAIARALVRPHIL 1193

QY 1200 LLDEATSAIDTESKVVQVEALDKAREGTCIVIAHRLSTIONADLIVVFONGVKHEGTH 1259
DB 1194 LLDEATSAIDTESKVVQVEALDKAREGTCIVIAHRLSTIONADLIVVIONGVKHEGTH 1253

QY 1260 QOLLAQKGIYFSMVSVQAGAKR 1281
DB 1254 QOLLAQKGIYFSMVSVQAGAKR 1275

RESULT 4
DVMS1
N:Alternate names: P-glycoprotein 1
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: A33719; A25057; I57510
R:Raymond, M.; Gros, P.
Proc. Natl. Acad. Sci. U.S.A. 86, 6488-6492, 1989
A>Title: Mammalian multidrug-resistance gene: correlation of exon organization with s
A:Reference number: A33719; MUID:89367274
A:Accession: A33719
A:Molecule type: DNA
A:Residues: 1-1276 <RAY>
R:Gros, P.; Croop, J.; Housman, D.
Cell 47, 371-380, 1986
A>Title: Mammalian multidrug resistance gene: complete cDNA sequence indicates strong
A:Reference number: A25057; MUID:87028229
A:Accession: A25057
A:Molecule type: mRNA
A:Residues: 1-1276 <GRO>
A:Cross-references: GB:M14757; NID:g199100; PIDN:AAA79005.1; PID:g387426
R:Raymond, M.; Gros, P.
Mol. Cell. Biol. 10, 6036-6040, 1990

A:Title: Cell-specific activity of cis-acting regulatory elements in the promoter of the
A:Reference number: I57510; MUID:91042535
A:Accession: I57510
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-21 <RNS>
A:Cross-references: GB:M60348; NID:q199102; PIDN:AAA39513.1; PID:g554199
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: mdr1 (pgpl)
A:Introns: 21/2; 38/3; 92/1; 111/2; 175/2; 233/3; 274/2; 332/3; 370/3; 407/3; 449/3; 517
A:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane pr
F:1-637,653-1276/Region: duplication
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:550-554/Region: nucleotide-binding motif B
F:1051-1247/Domain: ATP-binding cassette homology <ABC2>
F:1068-1075/Region: nucleotide-binding motif A (P-loop)
F:1194-1198/Region: nucleotide-binding motif B
F:73,91,96,103/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:432/Binding site: ATP (Lys) #status predicted
F:1074/Binding site: Atp (Lys) #status predicted

Query Match 81.9%; Score 5301; DB 1; Length 1276;
Best Local Similarity 80.3%; Pred. No. 2.1e-299;
Matches 1030; Conservative 130; Mismatches 114; Indels 8; Gaps 5;

QY 1 MDPEGGKSAENKFMKGGKSKKKEKPTVSTFAMFPRYSNWLDRMLVGTMAAII 60
DB 1 MEFEENLKGRADNFKSMGKSKK-EKKEKPAVGFGMEFYADWLKCMILGTAAII 59

QY 61 HGAALPLMLLVFGNMTDSFANAGISRNKTPVPIINSEITNNTQHFIN-HLEEMTYTAYY 119
DB 60 HGTLPLMLLVFGNMTDSFTKAES---ILPSITNQGPNSTLIISNSSLEEMTAYY 116

QY 120 YSGTGAGVLAAYIQVSWFGLAAGROLTKRKOFFHAIMROEIGWDFDHDVDELNRLTD 179
DB 117 YTGIGAGVLIVAYIQVSLWCLAGRLQHKIRQKFFHAIMQOEIGWDFDHDVDELNRLTD 176

QY 180 DVSKINEGIDKIGMFQSIATFTFTGIVTRGWLTLVLAISPLVGLSAAIAWAKILS 239
DB 177 DVSKINDIGDKIGMFQSIITFLAGFLIGFISGWLTLVLAISPLVGLSAAIAWAKVLT 236

QY 240 SFTDKELLAYAKAGAAEVLAIPTVIAFGGOKKELERYNKNLEAKGIGIKAITANI 299
DB 237 SFTNKLQAYAKAGAAEVLAIPTVIAFGGQOKELERYNKNLEAKNKGIGIKAITASI 296

QY 300 SIGAFLILYASALAFWYGTSLVLSSEYSGTGVLTFFSVLIGAFSIGOASPSIEAFAN 359
DB 297 SIGIAYLLVYASALAFWYGTSLVLSSEYSGTGVLTFFSVLIGAFSIGIHLAFNIEAFAN 356

QY 360 ARGAAFEIKIIDNKPSIDSYSGSKHPDNIKGNLEPKNVHFSYPSRKEYKILKGLNLKV 419
DB 357 ARGAAFEIKIIDNEPSIDSFSTGYRPSDSIMGNLEPKNVHFNYPSPSEVOILKGLNLKV 416

QY 420 QSGOTVALVNGSGCGKSTTVOLMORLYDPTDGMVCDIGQDIRINVRHLREITGVVSQEP 479
DB 417 KSGOTVALVNGSGCGKSTTVOLMORLYDPLEGVVSDIGQDIRINVRHLREITGVVSQEP 476

QY 480 VLFATTAENIRYRGENTVMDIEKAVKANAYDFIMKLPNKFDTLVGERGAOLSGGQK 539
DB 477 VLFATTAENIRYRGENTVMDIEKAVKANAYDFIMKLPNKFDTLVGERGAOLSGGQK 536

QY 540 RIARALVRNPKILLDDEATSDALDTESEAVVQVALDKARGRTTIVIAHRLSTVRNADY 599
DB 537 RIARALVRNPKILLDDEATSDALDTESEAVVQVALDKARGRTTIVIAHRLSTVRNADY 596

QY 600 IAGFDGIVYKGNHDELMEKGIYFKLVMTQTRGNEIELENATGESKSSDALEMSPKD 659
DB 597 IAGFDGIVYKGNHDELMEKGIYFKLVMTQTRGNEIELENATGESKSSDALEMSPKD 656

QY 660 SGSSLLIKRRSTRSIHAPQGDQRKLTGKEDLNENVPVSVFWRILKLKLNASTWPFYVVGIFC 719
DB 657 SKSPLI-RRSIYRSVHRKQDERLSMKEADEVDPVLSFWRILNLNLSEWYLLVGLVC 715

QY 720 AIINGLOPAFSIIFSRIGITFTRDEDEPETRQNSMPSVLVLGIIISFTIFFLQGFTE 779
DB 716 AVINGCIPQVFAIVSRVGVFSRDDHDETRQNCNLFSLFLVGLISFYTYFQGFTE 775

QY 780 GKAGELLKRLRYMVFWSMLRQDYSWFDDPKNTTGALTALTRLANDAAOVKAIGSRLAVIT 839
DB 776 KGAGELLKRLRYMVFWSMLRQDYSWFDDHKNSTGSLTRLASDASSVKGAMGARLAVT 835

QY 840 QNIANLGTGIIISLYGQWLTLILAIPIIAIAGVWEMKMLSGQALKKKELEGAGKIA 899
DB 836 QNVANLGTGIVLSLYGQWLTLILAVIPLVIGGIIEMKLLSGQALKKKELETSKIA 895

QY 900 TEAIENFRTVSLTREQKFEYMAQSLOVPTPNSLRKAHIFGVSEFISITQAMMYFSYACF 959
DB 896 TEAIENFRTIVSLTREQKFEYMAQSLOVPTPNSLRKAHIFGVSEFISITQAMMYFSYACF 955

QY 960 REGAVLVANEFMNFODVLLVFSIAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPL 1019
DB 956 REGAVLVAAQLMTFENVMLVFSVAVFGAMAGNTSSFPADYAKAKVSAHIIIEKTPE 1015

QY 1020 IDSYPHGLKPNLTGNTFNEVFNPTRPDIPVLOGLSLEVKGQTLALVSGSGCGKS 1079
DB 1016 IDSYTEGLKPTLLSGNVKFNQVFNPTRPNIPVLOGLSLEVKGQTLALVSGSGCGKS 1075

QY 1080 TVVQLLERYDPLAGSLVDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSIAENIAYGDS 1139
DB 1076 TVVQLLERYDPMAGSVFLDGKEIKQLNVQWLRHAHLGIVSQEPILFDCSIAENIAYGDS 1135

QY 1140 RVVSEELIQAANEIHHFIETLPEKYNTRVGDGTOLSGGOKORIAIARALVRQPHIL 1199
DB 1136 RAVSHEEIVRAAKEANIHOFIDSLDPKYNTRVGDGTOLSGGOKORIAIARALVRQPHIL 1195

QY 1200 LDEATSDALDTESEKVVQVQALDKAREGTCIVIAHRLSTIONADLIIVFQNGKVEHETH 1259
DB 1196 LDEATSDALDTESEKVVQVQALDKAREGTCIVIAHRLSTIONADLIIVFQNGKVEHETH 1255

QY 1260 QQLLAQGIYFMSVSVQAGAKR 1281
DB 1256 QQLLAQGIYFSM--VQAGAKR 1275

RESULT 5
JH0502
p-glycoprotein - rat
N:Alternate names: multidrug resistance protein mdr1b
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1992 #sequence revision 31-Mar-1992 #text_change 02-Feb-2001
C:Accession: JH0502; S22353; S22352
R:Silverman, J.A.; Raulio, H.; Gant, T.W.; Thorgeirsson, S.S.
Gene 106, 229-236, 1991
A:Title: Cloning and characterization of a member of the rat multidrug resistance (mdr)
A:Reference number: JH0502; MUID:92039081
A:Accession: JH0502
A:Molecule type: mRNA
A:Residues: 1-1277 <SIL>
A:Cross-references: GB:M62425
R:Deuchars, K.L.; Duthie, M.; Ling, V.
Biochim. Biophys. Acta 1130, 157-165, 1992
A:Title: Identification of distinct P-glycoprotein gene sequences in rat.
A:Reference number: S22351; MUID:92233089
A:Accession: S22353
A:Molecule type: DNA
A:Residues: 1212-1226, 'I', 1228-1277 <DEU>
A:Cross-references: EMBL:X61104; NID:g56890; PIDN:CAA43416.1; PID:g1334219
A:Accession: S22352
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1212-1226, 'I', 1228-1270, 'SV', 1271-1277 <DE2>
A:Cross-references: EMBL:X61103; NID:g56888; PIDN:CAA43415.1; PID:g1334218

C;Comment: This protein has the nucleotide binding motifs and ATP binding active transp
C;Gene: mdrlb
C;Superfamily: multidrug resistance protein; ATP-binding cassette homology
C;Keywords: ATP; duplication; glycoprotein; membrane protein; nucleotide binding; P-loop
F:409-603/Domain: ATP-binding cassette homology <ABC>
F:426-434/Region: nucleotide-binding motif A (P-loop)
F:530-554/Region: nucleotide-binding motif B
F:1053-1248/Domain: ATP-binding cassette homology <ABC2>
F:1070-1078/Region: nucleotide-binding motif A (P-loop)
F:1195-1199/Region: nucleotide-binding motif B
F:432/Binding site: ATP (Lys) #status predicted
F:1076/Binding site: ATP (Lys) #status predicted

Query Match 80.8%; Score 5231.5; DB 2; Length 1277;
Best Local Similarity 80.0%; Pred. No. 2.2e-295;
Matches 1029; Conservative 131; Mismatches 111; Indels 15; Gaps 9;

QY 1 MDPEGGRKSGAKNFWKMGKSKKNEKKKPTVSTFMPRYSNWLDRLYMLVGTMAAII 60
DB 1 MFEPEGNGRADKNFSKMGKSKK--EKEKKPAVIGFMFRYADWLDKLCMALGTAAII 58

QY 61 HGAALPLMLMLVFGNMTDFANAGTSRN-KTFPVLINESTNNTQHTN--HLEEMTTVA 117
DB 59 HGTLPLMLLVFGVYDTSFTP---SRDPHSDRAITNQSEINST-HVSDTSLEEDNAMA 114

QY 118 YYSGIGAGVLVAAIYQVSWCLAAAGROILKIRKQFFHAIMRQIEIGFDVHDVGEINTRL 177
DB 115 YYTGIGAGVLVAYIQVSWCLAAAGROIHKIROKFFHAIMNQIEIGFEDVNDAGELNTRL 174

QY 178 TDVSKINEGIDGKIGNFFOSIAFTFTGFTVTRGWKLTLVLALSPVLGLSAAIWKI 237
DB 175 TDVSKINDGIDGKLGFMFOISITFTSAGFTIGFISGWKLTTLVLAVSPVLGSSANMAKV 234

QY 238 LSSFTDKELLAYAKAGAAVEVLAAARTVIAFGQKKELERYNKNLEAKGIGIKKAITA 297
DB 235 LTSETNKELQAYAKAGAAVEVLAAARTVIAFGQKKELERYNKNLEAKRGVGIKAITA 294

QY 298 NISGAFLIYALAFWYGTSLVLSSEYSIGQVLTVPFVSIGVLTAFSIGQASPSIEAF 357
DB 295 NISGIAYLLVYASALAFWYGTSLVLSSEYSIGQVLTVPFVSIGVLTAFSIGHLAPNIEAF 354

QY 358 ANARGAAIEFKIIDNKPSIDTSKSGHKPDNIKLNKPNVHFSYPSRKEVKILKGLNL 417
DB 355 ANARGAAIEFKIIDNPSIDTSFTGHPKPSIDTSKSGHKPDNIKLNKPNVHFSYPSRKEVKILKGLNL 414

QY 418 KVSGGTVALVGNSSGCKSTTVQLMQLDPTDMVCMIDQDRTINVRHLREITGVWSQ 477
DB 415 KVSGGTVALVGNSSGCKSTTVQLMQLDPTDMVCMIDQDRTINVRHLREITGVWSQ 474

QY 478 EPLVFATTIAENTRYGRENVTWDEIEKAVKANAYDFIMKLPNKFTDLVGERGAQLSGQ 537
DB 475 EPLVFATTIAENTRYGRENVTWDEIEKAVKANAYDFIMKLPNKFTDLVGERGAQLSGQ 534

QY 538 KQRIATARLVNPKTKLLDEATSDLTSEAVQVALDKARKGRTTIVIAHRLSTVRNA 597
DB 535 KQRIATARLVNPKTKLLDEATSDLTSEAVQVALDKARKGRTTIVIAHRLSTVRNA 594

QY 598 DVTAGDDGVIVKGNHDELMKEGIFYFLVMTQTRGNEIELENATSKESDSEDALEMS 657
DB 595 DVTAGDDGVIVKGNHDELMKEGIFYFLVMTQTRGNEIELENATSKESDSEDALEMS 654

QY 658 KDSGSLIKRRSTRRSIHAPQODRKLGTKEDLNENPVPSFWRLKLNSTWEPYVVG 717
DB 655 EESKSPLI--RRSIRRSIHRRQDERRLSKEDVDVPMVSWQILKNISEWPYLVGV 713

QY 718 FCAILINGLQPAFSIIFSRIGFTREDPEPKRONSNMFVFLVLGLIISIFITFLOG 777
DB 714 LCAVINGLQPAFVIFSRIGFTREDPEPKRONSNMFVFLVLGLIISIFITFLOG 773

QY 778 TFGKAGEILTKRLRYMVFRSMLRQDVSWFDDPKNTTGATTLRLANDAAQVKAIGSR 837
DB 774 TFGKAGEILTKRLRYMVFRSMLRQDVSWFDDPKNTTGATTLRLANDAAQVKAIGSR 833

QY 838 ITONIANLGTGIIS--LIYQWQLTLLILAIIVPIIAAGVEMKMLSGOALKKKELEGA 895
DB 834 VTQNVANLGTGIISLVLYVYQWQLTLLILAIIVPIIAAGVEMKMLSGOALKKKELEIS 893

QY 896 GKIAEAIENFRTVSVSLTREQKFEYMYAQSLQVYPYNSLRKAHIFGVFSFISITQAMIFY 955
DB 894 GKIAEAIENFRTVSVSLTREQKFEYMYAQSLQVYPYNSLRKAHIFGVFSFISITQAMIFY 953

QY 956 ACDFREGALVLANEPNFDVLLVSAIVFGAMAGVQVSSFPADYAKAKVSAHVIMIE 1015
DB 954 AACDFREGALVLANEPNFDVLLVSAIVFGAMAGVQVSSFPADYAKAKVSAHVIMIE 1013

QY 1016 KSPILDSYSPHGLKPNLTGNTFNEVFNPTRPDIPVLOGLSLEVAKGOTLALVSGG 1075
DB 1014 KIPEIDSYSTEGKPNWLGNTFNGVKNFYTRNIPVLOGLSLEVAKGOTLALVSGG 1073

QY 1076 CGKSTVQVLLERFYDPLAGSLVDIGKEIKHLNVQWLRHAHLGIVSQEPILFDCSTAEINAY 1135
DB 1074 CGKSTVQVLLERFYDPLAGSLVDIGKEIKHLNVQWLRHAHLGIVSQEPILFDCSTAEINAY 1132

QY 1136 GNSRVWSHEETMOAKKANTHHFTETLPEKYNTRVGDGTQLSGGQKQRIATARALVRQ 1195
DB 1133 GNSRVWSHEETMOAKKANTHHFTETLPEKYNTRVGDGTQLSGGQKQRIATARALVRQ 1192

QY 1196 PHILLDEATSDTSEKVVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1255
DB 1193 PHILLDEATSDTSEKVVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1252

QY 1256 HGTHOOLLAQKGIYFSMVSVQAGAKR 1281
DB 1253 HGTHOOLLAQKGIYFSM--VOAGAKR 1276

RESULT 6

DWMSIA
Multidrug resistance protein la - mouse (fragment)
N:Alternate names: P-glycoprotein la
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: A34175
R:Hsu, S.I.H.; Lothstein, L.; Horwitz, S.B.
J. Biol. Chem. 264, 12053-12062, 1989
A:Title: Differential overexpression of three mdr gene family members in multidrug-re
A:Reference number: A34175; MUID:89308614
A:Accession: A34175
A:Molecule type: mRNA
A:Residues: 1-1104 <HSU>
A:Cross-references: GB:M24417; GB:J04839; NID:g200329; PIDN:AAA03243.1; PID:g200330
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c
structurally and functionally unrelated lipophilic antitumor drugs.
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane
F:234-428/Domain: ATP-binding cassette homology <ABC1>
F:251-258/Region: nucleotide-binding motif A (P-loop)
F:375-379/Region: nucleotide-binding motif B
F:877-1073/Domain: ATP-binding cassette homology <ABC2>
F:894-901/Region: nucleotide-binding motif A (P-loop)
F:1020-1024/Region: nucleotide-binding motif B
F:257/Binding site: ATP (Lys) #status predicted
F:900/Binding site: ATP (Lys) #status predicted

Query Match 77.2%; Score 4995; DB 1; Length 1104;
Best Local Similarity 89.3%; Pred. No. 9.6e-282;
Matches 985; Conservative 61; Mismatches 57; Indels 0; Gaps 0;

QY 179 DDVSKINEGIDGKTCMFEFQSTATFTFTGFTVTRGWKLTLVLALSPVLGLSAAIWKI 238
DB 1 DDVSKINEGIDGKTCMFEFQSTATFTFTGFTVTRGWKLTLVLALSPVLGLSAAIWKI 60

QY 239 SSFTDKELLAYAKAGAAVEVLAAARTVIAFGQKKELERYNKNLEAKGIGIKKAITAN 298
DB 61 SSFTDKELLAYAKAGAAVEVLAAARTVIAFGQKKELERYNKNLEAKGIGIKKAITAN 120

QY 174 NTRLTDDSVKINIGIGDKIGMPFQSTATFTGTFIVGFTRGWKLTLVLIAISPVGLSAAI 233
Db 174 NTRLTDDSVKINIGIGDKIGMPFQSTATFTGTFIVGFTRGWKLTLVLIAISPVGLSAAV 233
QY 234 WAKILSFTDKELLAYAKAGAAVEELAAIRTVIAFGGQKKELEYNKNLEAKGIGIKK 293
Db 234 WAKILSFTDKELLAYAKAGAAVEELAAIRTVIAFGGQKKELEYNKNLEAKGIGIKK 293
QY 294 AITANTISGAFLLIYASALAFWYGTSLVLSSEYISQGLVTVFFSVLIGAFSIGQASPS 353
Db 294 AITANTISGAFLLIYASALAFWYGTSLVLSSEYISQGLVTVFFSVLIGAFSIGQASPS 353
QY 354 IEAFANARGAAVEIPKIIDNKSIDSYKSGHKPNIDKGNLFEKNVHSPYSRKEVKILK 413
Db 354 IDAFANARGAAVEIPKIIDNKSIDSYKSGHKPNIDKGNLFEKNVHSPYSRANVKILK 413
QY 414 GLNLKVQSGOTVALVGNSSCGSTTVOLMORLYDPTDGMVCIQDGDRTINVRHLREITG 473
Db 414 GLNLKVQSGOTVALVGNSSCGSTTVOLMORLYDPTDGMVCIQDGDRTINVRHLREITG 473
QY 474 VVSOEPLVFTATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPKFDTLVGERGAQL 533
Db 474 VVSOEPLVFTATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPKFDTLVGERGAQL 533
QY 534 SGGOKORIAIARALVNPKNILLDEATSDALDESEAVVOALDKARKGRTTIVIAHRLST 593
Db 534 SGGOKORIAIARALVNPKNILLDEATSDALDESEAVVOALDKARKGRTTIVIAHRLST 593
QY 594 VRNADVIAGFDGVIYVEKGNHDELMEKEGKGYKPLVTMTQTRGMEIELEENATGSKSESAL 653
Db 594 VRNADVIAGFDGVIYVEKGNHDELMEKEGKGYKPLVTMTQTRGMEIELEENATGSKSESAL 653
QY 654 ESMPPKDGSSLLKRRSTRSIIAPAQQDRKLGTKED-LNENYPPVPSFWRILKLNSTEMPY 712
Db 654 ESMPPKDGSSLLKRRSTRSIIAPAQQDRKLGTKED-LNENYPPVPSFWRILKLNSTEMPY 712
QY 713 FVVGIFCAIINGLOPAFSIIIRIIGFTREDDEPTKQNSMFSVLVLVGIISFTIF 772
Db 713 FVVGIFCAIINGLOPAFSIIIRIIGFTREDDEPTKQNSMFSVLVLVGIISFTIF 772
QY 773 FLOGFTFGKAGETLLKRLRYMFRSMURDVSWEDEDPKNTTCALTRLANDAQAQVKGIG 832
Db 773 FLOGFTFGKAGETLLKRLRYMFRSMURDVSWEDEDPKNTTCALTRLANDAQAQVKGIG 832
QY 833 SRLAVITONIANLGTGIIISLYIYQWLTLLLIPIIPIAIAGVEMKMLSGOALDKKEL 892
Db 833 SRLAVITONIANLGTGIIISLYIYQWLTLLLIPIIPIAIAGVEMKMLSGOALDKKEL 892
QY 893 EGAGKIATEAENFRVTSLTREQPEYMYAOSLOVPYRNSLRKAHIFGVSFISQAMMY 952
Db 893 EGAGKIATEAENFRVTSLTREQPEYMYAOSLOVPYRNSLRKAHIFGVSFISQAMMY 952
QY 953 FSVAGCFRGAVLYVNGHMRFDVLVFSVIAVFGAVALGHASSFPADYAKAKLSAHLFM 1009
Db 953 FSVAGCFRGAVLYVNGHMRFDVLVFSVIAVFGAVALGHASSFPADYAKAKLSAHLFM 1009
QY 1013 IIEKSPIDSYSPHGLKPNLTGNTVFNENVNYPTRDPVLOGLSLEVKKGOFLALVG 1072
Db 1013 IIEKSPIDSYSPHGLKPNLTGNTVFNENVNYPTRDPVLOGLSLEVKKGOFLALVG 1072
QY 1073 SSGCGKSTVQVLLERFDPYDLAGSVLIDGKETHLNVQMLRHLGTVSQEPILFDCSIAEN 1132
Db 1073 SSGCGKSTVQVLLERFDPYDLAGSVLIDGKETHLNVQMLRHLGTVSQEPILFDCSIAEN 1132
QY 1133 IAYGDNRSRVVSHHEEIMORAKENIHHFTETLPEKYNTVRVGDKGTQSGQKQRTAIARAL 1192
Db 1133 IAYGDNRSRVVSHHEEIMORAKENIHHFTETLPEKYNTVRVGDKGTQSGQKQRTAIARAL 1192
QY 1193 VRQPHILLDEATSDALDESEKVVQEALDKAREGRTCTIVIAHRLSTIQNADLIYVVFQNG 1252
Db 1193 VRQPHILLDEATSDALDESEKVVQEALDKAREGRTCTIVIAHRLSTIQNADLIYVVFQNG 1252
QY 1253 VKEHGTHTOOLLAQKGIYFSMVSVOAGAK 1280

Db 1250 VKEHGTHTOOLLAQKGIYFSMVSVOAGTQ 1277
RESULT 8
DMS2
multidrug resistance protein 2 - mouse
N:Alternate names: P-glycoprotein MDR2
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: A30409; S70711
R:Gros, P.; Raymond, M.; Bell, J.; Housman, D.
Mol. Cell. Biol. 8, 2770-2778, 1988
A:Title: Cloning and characterization of a second member of the mouse mdr gene family
A:Reference number: A30409; MUID:88302195
A:Accession: A30409
A:Molecule type: mRNA
A:Residues: 1-1276 <HSU>
A:Cross-references: GB:J03398; NID:g199109; PIDN:AAA9516.1; PID:g387428
R:Kirschner, L.S.
Nucleic Acids Res. 24, 2829-2834, 1996
A:Title: De novo generation of simple sequence during gene amplification.
A:Reference number: S70711; MUID:96313253
A:Accession: S70711
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 43-92 <KLR>
A:Cross-references: EMBL:U46839; NID:g1228142; PIDN:AAC52722.1; PID:g1228143
C:Comment: The nucleotide sequence was submitted to the EMBL Data Library, January 1996
C:Note: This is an integral membrane protein overproduced in multidrug-resistant c
structurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: mdr2
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane
F:1-637,653-1276/Region: duplication
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:550-554/Region: nucleotide-binding motif B
F:1049-1245/Domain: ATP-binding cassette homology <ABC2>
F:1067-1074/Region: nucleotide-binding motif A (P-loop)
F:1192-1196/Region: nucleotide-binding motif B
F:88,94/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:432/Binding site: ATP (Lys) #status predicted
F:1072/Binding site: ATP (Lys) #status predicted
Query Match 75.1%; Score 4861; DB 1; Length 1276;
Best Local Similarity 74.8%; Pred. No. 7.1e-274;
Matches 960; Conservative 139; Mismatches 171; Indels 14; Gaps 8;
QY 1 MDPGGRRKSAEK--NFWKMGKSKK-NEKKKKPTVSTFAMFRYSNWLDRMLYMGTMA 57
Db 1 MDLEAARNGTARRLDGDFELGSGISNQGRRKKKVNKLIGLTLFRYSDWQDKLFWFGTLM 60
QY 58 AIITHGAALMLLVFGNMTDSFA-NAGISRNKTPFPVINESITNTQHTINHLEEEMTTY 116
Db 61 AIAHSGSLPLMWIVFGEMTKFDVNTG---NFSLPVNFSLSMLNPGR---ILEEEMTRY 113
QY 117 AYYSGTGAGVLVAAYIQVSFWCLAGRQLIKRKQFFHAIMRQEGWDFVHDVGEINTR 176
Db 114 AYYSGLGGVLAAYIQVSFWTLAGRQIKRKQFFHAILRQEMGWFIDIKGTTELNR 173
QY 177 LTDDVSKINGIGDKIGMFFQSIATFTFTGIVGTRGWKLTVLILATSPVLGLSAAIAWAK 236
Db 174 LTDDVSKISGIGDKVGNMFFQAIATATFAGFIVGIRGWKLTVLIMAIISPILGSTAVWAK 233
QY 237 ILSFTDKELLAYAKAGAAVEELAAIRTVIAFGGQKKELEYNKNLEAKGIGIKKAIT 296
Db 234 ILSTFSDKELAAVAKAGAAVEEAPALRTVIAFGGQKKELEYNKNLEAKGIGIKKALS 293
QY 297 ANTISGAFLLIYASALAFWYGTSLVLSSEYISQGLVTVFFSVLIGAFSIGQASPSIEA 356
Db 294 ANTISMGIAFLLIYASALAFWYGTSLVLSSEYISQGLVTVFFSVLIGAFSIGQASPSIEA 353

QY	357	PANARGAA	YEIFK	I	D	N	K	S	I	D	S	I	S	Y	S	K	S	G	H	K	P	N	I	K	G	N	L	E	F	K	N	V	H	S	P	S	R	K	V	K	I	L	G	L	N	416													
Db	354	PANARGAA	YVFI	O	I	D	N	N	K	I	D	S	F	S	R	G	H	K	P	N	I	K	G	N	L	E	F	S	D	V	H	S	P	S	R	A	N	I	K	L	G	L	N	413															
QY	417	LK	V	S	G	O	T	A	L	V	A	L	V	G	N	S	G	G	S	T	T	V	L	M	O	R	L	D	T	D	G	M	V	C	I	D	G	O	I	R	T	I	N	V	R	L	R	E	I	T	G	V	S	476					
Db	414	LK	V	S	G	O	T	A	L	V	A	L	V	G	N	S	G	G	S	T	T	V	L	L	O	R	L	D	T	D	T	E	G	K	I	S	I	D	G	O	I	R	N	F	N	V	R	L	R	E	I	T	G	V	S	473			
QY	477	Q	E	P	V	L	F	A	T	T	A	E	N	R	I	G	R	E	N	V	T	M	D	E	K	A	V	K	E	A	N	A	Y	D	F	I	M	K	L	P	N	K	F	D	T	L	V	G	E	R	G	A	O	L	S	G	536		
Db	474	Q	E	P	V	L	F	T	T	A	E	N	R	I	G	R	E	N	V	T	M	D	E	K	A	V	K	E	A	N	A	Y	D	F	I	M	K	L	P	Q	K	F	D	T	L	V	G	R	G	A	O	L	S	G	533				
QY	537	Q	K	O	R	I	A	I	A	R	A	L	V	R	N	P	K	I	L	L	D	E	A	T	S	A	L	D	T	E	S	A	V	Q	V	A	L	D	K	A	R	K	G	R	T	T	I	V	I	A	H	R	L	S	T	V	R	N	596
Db	534	Q	K	O	R	I	A	I	A	R	A	L	V	R	N	P	K	I	L	L	D	E	A	T	S	A	L	D	T	E	S	A	V	Q	A	L	D	K	A	R	E	G	R	T	T	I	V	I	A	H	R	L	S	T	V	R	N	593	
QY	597	A	D	V	I	A	G	P	D	G	V	I	E	K	N	H	D	E	L	M	K	E	G	I	Y	P	K	L	V	T	M	O	T	R	N	E	T	E	L	E	N	A	T	E	S	E	S	D	A	L	E	M	S	656					
Db	594	A	D	V	I	A	G	P	D	G	V	I	E	Q	S	H	S	E	L	M	K	E	G	I	Y	P	R	L	V	N	M	O	T	A	G	S	I	L	S	E	F	E	L	S	E	D	E	K	A	A	G	D	653						
QY	657	P	K	D	S	G	S	L	I	K	R	R	T	S	I	H	A	P	O	D	R	K	T	G	E	D	L	A	N	E	N	P	P	S	E	F	W	I	L	K	N	T	E	M	P	F	V	G	716										
Db	654	P	-	N	G	W	A	R	I	P	R	N	T	K	S	L	K	S	P	H	-	O	N	R	L	D	E	T	N	E	L	D	A	N	P	P	S	F	L	V	K	L	N	T	E	M	P	F	V	G	711								
QY	717	I	F	C	A	I	N	G	G	I	O	P	A	F	S	I	I	F	R	I	G	I	F	T	R	D	E	P	T	K	R	Q	N	S	M	F	S	V	L	F	V	L	G	I	I	S	F	T	F	F	L	Q	776						
Db	712	T	V	C	A	I	N	G	A	L	O	P	A	F	S	I	I	S	E	M	A	I	F	G	P	D	-	A	V	K	O	K	C	N	M	F	S	V	L	P	L	G	I	V	L	S	F	T	F	F	L	Q	770						
QY	777	F	T	F	G	K	A	G	E	I	L	T	K	L	R	Y	M	V	R	S	M	L</																																					

RESULT 9

I48123

p-glycoprotein isoform III - Chinese hamster

C:Species: *Cricetulus griseus* (Chinese hamster)

C,Species: *Citellus griseus* (Chinese hamster)
C,Date: 02-Jul-1996 #sequence revision 02-Jul-1996 #text change 02-Feb-2001

C:Accession: I48123

R;Endicott, J.A.; Sarangi, F.: Ling, V.
C/ACCESSION. 140123

DNA Seq. 2, 89-101, 1991
A:Title: Complete cDNA sequences encoding the Chinese hamster P-glycoprotein gene fam
A:Reference number: I48121; MUID:92135896
A:Accession: I48123
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1281 <RES>
A:Cross-references: GB:M60042; NID:g191168; PIDN:AA58885.1; PID:g191169
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop
F:412-606/Domain: ATP-binding cassette homology <ABC1>
F:429-436/Region: nucleotide-binding motif A (P-loop)
F:1054-1250/Domain: ATP-binding cassette homology <ABC2>
F:1071-1078/Region: nucleotide-binding motif A (P-loop)

Query Match 74.5%; Score 4823.5; DB 2; Length 1281;
Best Local Similarity 73.7%; Pred. No. 1.1e-271;
Matches 949; Conservative 147; Mismatches 175; Indels 17; Gaps 8;

Db 772 FLQFTFGKAGEILTRRLSRMAFKAMLRQDMWFDYDKNSTGALSTRLATDRAQVQATG 831
QY 833 SRLAVITQNTANLGTGIIISLYIGWQLTLLLAIVPIAIAAGVEMKMLSGQALKDKKL 892
Db 832 TRALIAQTANLGTGIIISLYIGWQLTLLLSVVPFIAVSGIVEMKMLAGNAKDKKAL 891
QY 893 EGAGKATEAIENFRVWVSLTREQFEYMAQSLQVPVYNSLRKAHIFGVSPSIQAMMY 952
Db 892 EAAGKATEAIENIRVVSITQERFESMYVEKLHEPYNSVQMAHIYGFISISQAFMY 951
QY 953 FSVAGCFRGAYLVANEFMNFODVLVFSIAIVGAMAVQVSSFPADYAKAKVSAHVIM 1012
Db 952 FSVAGCFRGAYLVNHRMFRDVLVFSIAIVGAVGALGHASSFADYAKAKLSAHLFS 1011
QY 1013 IIEKSLIDSYSPHGLKPNLTGNTFNENYVNYTRPDIPVLOGLSLEVKKGTALVG 1072
Db 1012 LFERQPLIDSYSGEGLWPKDFEGSVTFNEVFNYPTRANMPVLGSLLEVKKGTALVG 1071
QY 1073 SSGCGKSTVQLLERYDPLAGSLVDGKEIKHLNVQMLRAHLGIVSQEPILFDCSIAEN 1132
Db 1072 SSGCGKSTVQLLERYDPMAGTVLLDGOEAKKLNIQWLRAQLGIVSQEPVLFDCSIAEN 1131
QY 1133 IAYGDSNRVSVHEEINQAAKANIHHFIETLPKYNTRVGDGKTQSGQKORIAIARAL 1192
Db 1132 IAYGDSNRVSVDEIVRAAKAANIHFIEPLPKYKTRVGDGKTQSGQKORIAIRAL 1191
QY 1193 VROPHILLDEATSDLTSEKVVQALDKAREGRTCIIVIAHRLSTIONADLIVVFQNGK 1252
Db 1192 IROPRVLLDEATSDLTSEKVVQALDKAREGRTCIIVIAHRLSTIQNADLIVVQNGK 1251
QY 1253 VKEHGHTHQQLAQKGIYFSMVSVQAGK 1280
Db 1252 VKEHGHTHQQLAQKGIYFSMVNIQAGK 1279

RESULT 10
S41646
P-glycoprotein - rat
A:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 02-Feb-2001
C:Accession: S41646; S22354
R:Brown, P.C.; Thorgerlsson, S.S.; Silverman, J.A.
Nucleic Acids Res. 21, 3885-3891, 1993
A:Title: Cloning and regulation of the rat mdr2 gene.
A:Reference number: S41646; MUID: 93376516
A:Accession: S41646
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-1278

A:Cross-references: EMBL:L15079; NID:g310192; PIDN:AAA02937.1; PID:g310193
R:Deuchars, K.L.; Duthie, M.; Ling, V.
Biochim. Biophys. Acta 1130, 157-165, 1992
A:Title: Identification of distinct P-glycoprotein gene sequences in rat.
A:Reference number: S22351; MUID: 92223089
A:Accession: S22354
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1211-1278 <DEU>
A:Cross-references: EMBL:X61105; NID:g56892; PIDN:CAA43417.1; PID:g1334220
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:1051-1247/Domain: ATP-binding cassette homology <ABC2>
F:1068-1075/Region: nucleotide-binding motif A (P-loop)

Query Match 74.3%; Score 4808; DB 2; Length 1278;
Best Local Similarity 73.9%; Pred. No. 8.5e-271;
Matches 949; Conservative 146; Mismatches 176; Indels 14; Gaps 8;

QY 1 MDPEGRKGSAAK-NFWKMGKSKKN-EKKEKKPTVSTFAMFRYSNWLDRMLVGTMA 57
Db 1 MDLEAARNGTARRLDGDFELGSISSNQSKKKKKNLIGPLTLFRYSWQDKLFLMLGTAM 60

QY 58 AIIHGAALPLMLYFGNMTDSFA-NAGISRNKTFPVIINESITNNTQHFINHLEEMTTY 116
Db 61 AIAHSGSLPLMWIVGEMTDKFDVNDAG--NFSLPVNFSLMNLNPGR---ILEEMTRY 113
QY 117 AYYSGIGAGVLVAAVYQVSWFCLAAAGQILKIRKQFFHAIMRQEIOWFDVHDVGENTR 176
Db 114 AYYSGIGAGVLVAAVYQVSWFCLAAAGQILKIRKQFFHAIMRQEIOWFDVHDVGENTR 173
QY 177 LTDDVSKINEGIGDKIGMFTQSIATFTFTGTVGTGKWLTVLALSPVLGISAALWAK 236
Db 174 LTDDVSKINEGIGDKIGMFTQSIATFTFTGTVGTGKWLTVLALSPVLGISAALWAK 233
QY 237 ILSFTDKELLAYAKAGAVAEVLAARTVIAFGGQKKELERYNKNLEAEKIGIKKAIT 296
Db 234 ILSFTDKELLAYAKAGAVAEVLAARTVIAFGGQKKELERYNKNLEAEKIGIKKAIT 293
QY 297 ANISGAFLLIYASIALAFWYGTSLVLSSEYSIGQVLTVFFSVLIGAFSIGAQSPIEA 356
Db 294 ANISGAFLLIYASIALAFWYGTSLVLSSEYSIGQVLTVFFSVLIGAFSIGAQSPIEA 353
QY 357 FANARGAAEYEFKIIDNKPSIDSYSKSHKPDNLKLEFNHVSYPSPKVKILKGLN 416
Db 354 FANARGAAEYEFKIIDNKPSIDSYSKSHKPDNLKLEFNHVSYPSPKVKILKGLN 413
QY 417 LKVGSGQTVLVGNSGCGKSTTVQLMQRLYDPTDMVCIIDQDITINVRHLREITGVVS 476
Db 414 LKVGSGQTVLVGNSGCGKSTTVQLMQRLYDPTDMVCIIDQDITINVRHLREITGVVS 473
QY 477 QEPVLFTATTAEINRYGRENVTMDIEKAVKEANAYDFIMKLPKFTIVGERGAQLSGG 536
Db 474 QEPVLFTATTAEINRYGRENVTMDIEKAVKEANAYDFIMKLPKFTIVGERGAQLSGG 533
QY 537 QKQRIATARALVRNPKILLDEATSDLTSEAEVQALDKAREGRTTIIVIAHRLSTVRN 596
Db 534 QKQRIATARALVRNPKILLDEATSDLTSEAEVQALDKAREGRTTIIVIAHRLSTVRN 593
QY 597 ADVIAGFDGIVIVEKGNHDELMEKGIYFKLVTMTQNRGNEIELEENATGESKESDALEMS 656
Db 594 ADVIAGFDGIVIVEKGNHDELMEKGIYFKLVTMTQNRGNEIELEENATGESKESDALEMS 653
QY 657 PKDSSGLIKRRSTRSIHAPQGDQDKLGTG-EDLNENVPVPSFWRLKLNSTWEPYFVV 715
Db 654 P-NGWKARIFRNSTKSLKSRHQNLVDVETNELDANVPVPSFWRLKLNSTWEPYFVV 712
QY 716 GIFCAITNGGLQPAFISIFSRIGIFTRDEDPETKRONSNFVSLFLVLGIGIFITFLQ 775
Db 713 GIFCAITNGGLQPAFISIFSRIGIFTRDEDPETKRONSNFVSLFLVLGIGIFITFLQ 771
QY 776 GFTFGKAGEILTKRLRYNFRSMRLQDVSWFDDPKNTTGALTTRLANDAAQVGAIGSRL 835
Db 772 GFTFGKAGEILTKRLRYNFRSMRLQDVSWFDDPKNTTGALTTRLANDAAQVGAIGSRL 831
QY 836 AVTQNTANLGTGIIISLYIGWQLTLLLAIVPIAIAAGVEMKMLSGQALKDKKELEGA 895
Db 832 AVTQNTANLGTGIIISLYIGWQLTLLLAIVPIAIAAGVEMKMLSGQALKDKKELEGA 891
QY 896 GKIAEAIENIRTVVSLTREQFEYMAQSLQVPVYNSLRKAHIFGVSPSIQAMMYFSY 955
Db 892 GKIAEAIENIRTVVSLTREQFEYMAQSLQVPVYNSLRKAHIFGVSPSIQAMMYFSY 951
QY 956 AGCFRFGAYLVANEFMNFQDVLVFSIAIVGAMAVQVSSFPADYAKAKVSAHVIMIE 1015
Db 952 AGCFRFGAYLVANEFMNFQDVLVFSIAIVGAMAVQVSSFPADYAKAKVSAHVIMIE 1011
QY 1016 KSPIDSYSPHGLKPNLTGNTFNENYVNYTRPDIPVLOGLSLEVKKGTALVGS 1075
Db 1012 KSPIDSYSPHGLKPNLTGNTFNENYVNYTRPDIPVLOGLSLEVKKGTALVGS 1071
QY 1076 CGKSTVQLLERYDPLAGSLVDGKEIKHLNVQMLRAHLGIVSQEPILFDCSIAENIAY 1135
Db 1072 CGKSTVQLLERYDPLAGSLVDGKEIKHLNVQMLRAHLGIVSQEPILFDCSIAENIAY 1131

QY 1136 GDSRVVSHHEIMQAKEANIHHFIETLPKYNTRVGDKTQVLSGGQKQRIATARALVRO 1195
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 1132 GDSRVVSDQEIIVRAAKEANIHPETLPQKRYETRVGDKTQVLSGGQKQRIATARALRQ 1191
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 1196 PHILLDEATSEKVVQVQALDKAREGRTTIVIAHRLSTONADLIVVFQNGKVKRE 1255
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 1192 PRVLLDEATSEKVVQVQALDKAREGRTTIVIAHRLSTONADLIVVINDGKVKRE 1251
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 1256 HGTHQQLLAQKGIYFSMVSVQAGAK 1280
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 1252 HGTHQQLLAQKGIYFSMVNIQAGTQ 1276
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||

RESULT 11
S55692
multidrug resistance protein homolog (mdr) - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 02-Feb-2001
C:Accession: S55692
R:Castillo, G.; Shen, H.J.; Horwitz, S.B.
Biochim. Biophys. Acta 1262, 113-123, 1995
A:Title: A homologue of the mammalian multidrug resistance gene (mdr) is functionally ex
A:Reference number: S55692; MUID:95322451
A:Accession: S55692
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1287 <C>
A>Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; nucleotide binding; P-loop
F:420-614/Domain: ATP-binding cassette homology <ABC1>
F:437-444/Region: nucleotide-binding motif A (P-loop)
F:1062-1258/Domain: ATP-binding cassette homology <ABC2>
F:1079-1086/Region: nucleotide-binding motif A (P-loop)

Query Match 68.4%; Score 4427; DB 2; Length 1287;
Best Local Similarity 67.6%; Pred. No. 1.1e-248;
Matches 871; Conservative 172; Mismatches 211; Indels 34; Gaps 11;

QY 2 DPEGRKGSABKFWKGGKSKKKEKKKPTVSTFAMFRYSNWLDRILVMTGMAIHH 61
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 21 DPNSK--EKKGFF--SKFKKKKEKTEKPKVGVTFTFRYSSTSDKMLMFGTIAHLAH 76
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 62 GAALPLMLVFGNMTDSPANAG--ISRNKTPPVIIINESITNTQHFNHLEEMTTVAYY 119
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 77 GAALPLMLVFGNMTDSPANAG--ISRNKTPPVIIINESITNTQHFNHLEEMTTVAYY 127
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 120 YSGICAGVLVAAIYQVSWCLAAQRILKIRKQFFHAIMRQEIQWFDVHDVGLNLTLD 179
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 128 YSGLGFGVMLCAYIQISFWLTSAGQIKKIRSNFFHVLVQEIQWFDINDAGELNLTLD 187
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 180 DVSKINEGIGKIGMFQSIATFTTGTIVGTGTRGKWLTLVLTAISPVGLSAAIWAHLS 239
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 188 DVSKINEGIGKIAMLLQSLTTLTGTFTIGIKGWLTLVWGAISPIMGLSAAIWAHLS 247
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 240 SFTDKELLAYAKAGAAVEVLAAITVIAFGQKKELERYNKNLEAKGIGIKKAITANI 299
:||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 248 AFTNKLKAYAKAGAAVEVLSSITVFAFGQKNEIHRYEKNLEDAKKIGIKKAITANY 307
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 300 SIGAFLIYASALAFWYGTSLVLSSEYISGQVLTFFVFSVLIGAFSIGQASPIEAFAN 359
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 308 SIGFAFLIYASALAFWYGTSLVLSSEYISGQVLTFFVFSVLIGAFSIGQASPIEAFAN 367
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 360 ARGAAIETFKIDNKPISDYSKSGHKPDNTKGNLEPKNVHFSYPSRKEVKILKGLNLKV 419
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 368 ARGAAITFNIDNPKIDSFSGKGLPKDKIGDIEFKNVITFTYPSRKDIQVLKGLNLNI 427
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 420 QSGQTVLVGSGCGKSTVQIMORLYDPTDGMVICDQDITINVRHLREITGVVSOEP 479
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 428 PSQKTVLVGSGCGKSTVQIMORLYDPTDGMVICDQDITINVRHLREITGVVSOEP 487
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 480 VLFATTIAENTRYGRENVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQK 539
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||

Db 488 ILFDTTIADNIRYGRDVTKEIERATKEANAYDFIMKLPDKLETIVLGERGTQLSGGQK 547
QY 540 RTAIALRVNPKILLDDATSEAVQVQALDKAREGRTTIVIAHRLSTVYRNADY 599
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 548 RTAIALRVNPKILLDDATSEAVQVQALDKAREGRTTIVIAHRLSTVYRNADY 607
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 600 IAGFDGVIIVEKGNHDELMEKGIYFKVLTMTQ---RGNEIELEENATGESKESDALEM 655
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 608 IAGFDGVIIVEKGNHDELMEKGIYFKVLTMTQ---RGNEIELEENATGESKESDALEM 660
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 656 SPKDSGSSILKRRSTRSIIHA--POQDRKLGTKE--DLNENVPVPSFWIRILKLNTEWPPY 712
: ||:||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 661 PVTHTSNLVRKSSNTIKSPVETDEKVEDEEKKKEGPPVPSFFKVMKLNPEWPPY 720
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 713 FVVGIFCAIINGLOPAFSIIFSRIGIETREDDEPK--RQNSNMFSVLVLVLGIISFT 771
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 721 FVVGIVCAIINGLOPAFSIIFSRIGIETREDDEPK--RQNSNMFSVLVLVLGIISFT 777
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 772 FFLQGTFFGKAGIILTKRLRYMVFRSMLRQDYSWDFDPKNTTGALTTLRLANDAAQVKGAI 831
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 778 FFLQGTFFGKAGIILTKRLRYMVFRSMLRQDYSWDFDPKNTTGALTTLRLANDAAQVKGAI 837
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 832 GSRILAVITONIANLGTGIIISLIYQWQLLLALLAIVPIITAIAGVVMKMLSGQALKDKKE 891
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 838 GTRALLAQNVAALGTIIISLIYQWQLLLALLAIVPIITAIAGVVMKMLSGQALKDKKE 897
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 892 LEGAGIATEIENFTVVSLETRQKFEYVYAGSLQVYRNSILRKAHIFGVSFSTQAMM 951
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 898 LEKAGIKISTDAVINITVVSLETRQKFEYVYAGSLQVYRNSILRKAHIFGVSFSTQAMM 957
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 952 YFSYAGCFR--FGAYLVANBFMFQDVLVLSAIVFGAMAVGVSSFPADYAKAKYSAAHV 1010
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 958 VLCLWVSVGLAYLVVEGLMKLDEVLYSSAIVLGMALGQTSFPADYAKAKYSAAHV 1017
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 1011 IMIERSPLIDSYPHGLKPNLTLEGWNTVNEVFNTPRDPIDVLOGLSLEKVKGGTIAL 1070
: ||:||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 1018 FSLERVPQIDSYSDQGEKPKNGSNVFPKGNFNPTRPDITVLOGLDYSVKQGETLAL 1077
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 1071 VGSAGCGKSVVOLLERFDPLAGSVLIDGKEIKHLNVQWLAHLGIVSQEPILFDCSTA 1130
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 1078 VGSAGCGKSVVOLLERFDPLAGSVLIDGKEIKHLNVQWLAHLGIVSQEPILFDCSTA 1137
||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 1131 ENIAYGDNRSVSHHEIMQAKEANIHHFIETLPKYNTRVGDKTQVLSGGQKQRIATAR 1190
:||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 1138 DNIAYGDNRSVSHHEIMQAKEANIHHFIETLPKYNTRVGDKTQVLSGGQKQRIATAR 1197
:||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 1191 ALVQPHILLDDATSEKVVQVQALDKAREGRTTIVIAHRLSTONADLIVVFQNGKVKRE 1255
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 1198 ALVQPHILLDDATSEKVVQVQALDKAREGRTTIVIAHRLSTONADLIVVFQNGKVKRE 1257
| :||||||| :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
QY 1251 GKVEGHTHOQLLAQKGIYFSMVSVQAG 1278
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||
Db 1258 GKVEGHTHOQLLAQKGIYFSMVSVQAG 1285
||| | :||:||||||| :||||||| :||||||| :||||||| :||||||| :||||||| :||

RESULT 12
T42228
P-glycoprotein sister - rat
N:Alternate names: multidrug resistance protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Feb-2001
C:Accession: T42228
R:Childs, S.J.; Yeh, R.L.; Hui, D.; Ling, V.
submitted to the EMBL Data Library, June 1997
A:Description: Taxol resistance mediated by the liver-specific sister gene of P-glyco
A:Reference number: 222088
A:Accession: T42228
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1321 <CH>
A:Cross-references: EMBL:AF010597; NID:g3273483; PID:g3273484; PIDN:AAC24753.1
C:Genetics:
A:Gene: spgp

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C;Keywords: glycoprotein; membrane protein; nucleotide binding; P-loop
F;455-462/Region: nucleotide-binding motif A (P-loop)

Query Match 49.7%; Score 3214.5; DB 2; Length 1321;
Best Local Similarity 50.4%; Pred. No. 2.6e-178;
Matches 649; Conservative 233; Mismatches 370; Indels 35; Gaps 11;

Qy 20 KSKKNEK-KEKKPTVSTFAMFRYSNWLDRLVLMVCTMAAITHGAALPLMLVFGNMTDS 78
Db 30 KKSRLQDKMKEGDIRVGFELFRSSSKDIWLMGMGVCALLHGMAGPGLIIFGIMTDI 89
Qy 79 FANAGISR-----NKTPEPVINES-----ITNNTQHFHNLEEMTTVAYYSGI 123
Db 90 FIKYDIERQELIPEKACVNTI-VWINSFQNMNTGTVCGLVDIESEMIKFSGIYAGV 148
Qy 124 GAGVLVAAYIQVSWFCLAAQROILKIRKQFFHAIMRQEGFDVHDVGNELTRLTDVSK 183
Db 149 GMTVLILGYFOIRLWITGARQIRRMKIFYRRIMRMEIGWFDCTSVGELNSRFADDIEK 208
Qy 184 INEGIGDKIGMFFQSIAATFTGTFVGRGWKLTVLILAIISPVGLSAAIAWAKILSSFTD 243
Db 209 INDIAIOLAHLFQRMSTAMCGLLLGIFYRGWKLTLVLAVSPILIGIAVIGLSIAKFT 268
Qy 244 KELLAYAKAGABEVLAAIIRTVIAFGQKKELERYKNLEAKGIGIKKAITANISIGA 303
Db 269 LELKAYAKAGSIADVLSSIRTVAAPGGENKEVEREKNLVFAQRWIKGMVMGFETGY 328
Qy 304 AFLLIYASALAFWYGTSLVL-SSEYSIGQVLTVPFSLVIGAFSGQASPSIEAFANARG 362
Db 329 MWCLIFFCYALAFWYGTSLVLDEEYTPGTVLQIFCLVLAAMNIGHASSCLEIFSTGCS 388
Qy 363 AAYEIFKIIDNPKSIDSKGHKPDNIKGNLEFNHVSYPSEKVKILKGLNKLKVGSG 422
Db 389 AATNIFQTDIQPVDCMSDGYKLDRIKEIEFHNVTHEYSPDPVKILDLNLSWIKPG 448
Qy 423 QTVALVNSCGCKSTVTQVMQRLYDPTDGMVCDIGODIRTNVRLHREITGVVSQEPVL 482
Db 449 ETTALVSGSGAGKSTALQIIFYDPCGVTLDGHDIRSLNRLWDQIGIVEQEPVL 508
Qy 483 ATTIAENIRYGRNVTMDIEKAVKANAYDFTMKLPKFDPLVGERGALSGQKQRTA 542
Db 509 STTIAENIRFGRDATMEDIVQAQDANAYNFIMALPOQFDPLVGEVGGGMSGQKQV 568
Qy 543 IARALRNPKILLDLDPATSDATESAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAG 602
Db 569 IARALRNPKILLDMATSDALDNESEARVQEAALNKHQHTTIISVAHRLSTVRADVI 628
Qy 603 FDCGVIVEGNHDELMKEGYFKLVMTQTRNEIELENA-TGESKSESDALEM-----S 656
Db 629 FEHGVAVERTGHEELLERKGVFMVLVTLQSGDNHAKETSINGKDATEGGTLERTFSR 688
Qy 657 PKDSGSLIKRSTR--SIHAP--QGDRKLGTEK-----LNENVPYVFWILKIN 706
Db 689 YRDSLRSIRQSKSLSLTHDPLAVADHKSYSKSDNDVLVEEYEPAPVRILKYN 748
Qy 707 STEWPYVVGIFCAINGLOPAFSIIFSRIGIETREDDEPKTKQNSMFSVFLVLGI 766
Db 749 IPBWHVILVGLSAAINGAVTPIYSLLFSQLLGTSL-LDKEQQSEIHSNCLFFVILGC 807
Qy 767 ISFITFLOGFTGKAGEILTRRLRYMFRSMRLQDVSWFDDPKNTGALTTRLANDAAQ 826
Db 808 VSIPTFLOGYTFKASGELLTRLRKFGFKAMLGQIDGWDFDLRNNPGLVTLRLTADASQ 867
Qy 827 VKGAGSRLAVTONTANTGTGIIISLIYGMQTLTLLLAIVPIIAIAGVVMKMLSGAL 886
Db 868 VQATGSGVMGNVSTNIIAALLTAFTFSWKLSLIITFFPFLSAGAVOTKMLTGPAS 927
Qy 887 KDKKELEGAKIATEAENFRVWLSLREQKPEYMAQSLQVPYRNSRKAHIFGVSPSI 946
Db 928 QDKQALEKAGQITSEALSNIRTVAGIGVEGRIFAKAFVELQTSYKATVRKANIYGLCF 987
Qy 947 TQAMMYFVAGCFRFGALVLANEFMNFQDVLVFSALVFGAMAVGVQVSSFADYAKAKVS 1006

Db 988 SQGIAFLANSAAARYGGVGLIAVEGLGFSHVFRVSSVLSATAVGRFTSYTPSYAKAKIS 1047
Qy 1007 AAHVIMITEKSPDLDSYSPHGLKPNLTLEGNTFNEVFNYPTRDPDIPVLQGLSLEVKKQ 1066
Db 1048 AARFQLDLDRKPPINVINSEAGEKWNDFQKIDFIDCKFTYPSRDPDIQVLNGLSVNPGQ 1107
Qy 1067 TLALVSSGCGCKSTVQQLLERYDPLAGSLVDGKEIKHLNVQWLRHGLVTSQEPILFD 1126
Db 1108 TLAIVGSSGCGCKSTQILERYDQDGTVMIDGHSKKVNIQILRSNIGVSGEPVLFD 1167
Qy 1127 CSIAENIAYGNSRVSRVSHHEEIMQAANEIHHFIETLPEKYNTNRVGDKGTSOLSGQKRI 1186
Db 1168 CSIMDNIKYGDNTREISVERAIAAAQAQLHDFVMSLPEKYETNVGIGQSLSRGEKQRI 1227
Qy 1187 AIARALVROPHILLDEATSDLTSEKVVQEAALDKAREGRTCIVIAHRLSTIONADLIV 1246
Db 1228 AIARAIVRDPKILLDEATSDLTSEKVTQALDKAREGRTCIVIAHRLSTIONSDIIA 1287
Qy 1247 VFQNGKVKHGTQOOLLAOKGIYFSMV 1273
Db 1288 VVSQGVVIEKGTHEKLMQAOKGAYYKLV 1314

RESULT 13

T42842

N;Alternate names: bile salt export protein, ATP-dependent - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-May-2000

C;Accession: T42842

R;Gerloff, T.; Stieger, B.; Hagenbuch, B.; Madon, J.; Landmann, L.; Roth, J.; Hofmann

J. Biol. Chem. 273, 10046-10050, 1998

A;Title: The sister of P-glycoprotein represents the canalicular bile salt export pum

A;Reference number: Z22272; MUID:98212048

A;Accession: T42842

A;Status: preliminary;

A;Molecule type: mRNA

A;Residues: 1-1321 <GER>

A;Cross-references: EMBL:U69487; NID:g3075421; PID:g3075422; PIDN:AAC40084.1

A;Experimental source: strain Sprague-Dawley; liver

C;Gene: spgp

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

C;Keywords: Atp; duplication; glycoprotein; membrane protein; P-loop

Query Match 49.6%; Score 3210.5; DB 2; Length 1321;
Best Local Similarity 50.3%; Pred. No. 4.9e-178;
Matches 648; Conservative 233; Mismatches 371; Indels 35; Gaps 11;

Qy 20 KSKKNEK-KEKKPTVSTFAMFRYSNWLDRLVLMVGTMAAIIHGAALPLMLVFGNMTDS 78
Db 30 KKSRLQDKMKEGDIRVGFELFRSSSKDIWLMGMGVCALLHGMAGPGLIIFGIMTDI 89
Qy 79 FANAGISR-----NKTPEPVINES-----ITNNTQHFHNLEEMTTVAYYSGI 123
Db 90 FIKYDIERQELIPEKACVNTI-VWINSFQNMNTGTVCGLVDIESEMIKFSGIYAGV 148
Qy 124 GAGVLVAAYIQVSWFCLAAQROILKIRKQFFHAIMRQEGFDVHDVGNELTRLTDVSK 183
Db 149 GMTVLILGYFOIRLWITGARQIRRMKIFYRRIMRMEIGWFDCTSVGELNSRFADDIEK 208
Qy 184 INEGIGDKIGMFFQSIAATFTGTFVGRGWKLTVLILAIISPVGLSAAIAWAKILSSFTD 243
Db 209 INDIAIOLAHLFQRMSTAMCGLLLGIFYRGWKLTLVLAVSPILIGIAVIGLSIAKFT 268
Qy 244 KELLAYAKAGABEVLAAIIRTVIAFGQKKELERYKNLEAKGIGIKKAITANISIGA 303
Db 269 LELKAYAKAGSIADVLSSIRTVAAPGGENKEVEREKNLVFAQRWIKGMVMGFETGY 328
Qy 304 AFLLIYASALAFWYGTSLVL-SSEYSIGQVLTVPFSLVIGAFSGQASPSIEAFANARG 362
Db 329 MWCLIFFCYALAFWYGTSLVLDEEYTPGTVLQIFCLVLAAMNIGHASSCLEIFSTGCS 388

Db 1118 SLREQVICVSOEPTIFDCTIGENICYGTN-RNVTYQEIVEAAKMANIHNFILGLPDGYDT 1176
Qy 1170 RVGDKGTOLSGGQKORIAIARALYRPHILLDEATSAIDTSEKVVQVQALDKAREGRTC 1229
Db 1177 HVGEGTOLSGGQKORIAIARALYRSPVLLDEATSAIDTSEKIVQVQALDAAKQGRTC 1236
Qy 1230 IVIAHRLSTIONADLIVVFQNGKVEHGT HQOILLAQRGIYFSMVSQ 1276
Db 1237 LVIAHRLSTIONSVDIAIVSEKIVEKGT HDELIRKSEIYQKFCETQ 1283

Search completed: November 6, 2002, 18:44:49
Job time : 27.063 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:30:44 : Search time 10.4981 Seconds
(without alignments)
4724.646 Million cell updates/sec

Title: US-09-672-725C-23

Perfect score: 6474

Sequence: 1 MDPEGRKSGSAENFWKMGK.....LLAQKGIYFSMVQAGAKR 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5866	90.6	1280	MDR1_HUMAN	P08183 homo sapien
2	5667.5	87.5	1276	MDR1_CRIGR	P21448 cricetus
3	5638	87.1	1276	MDR3_MOUSE	P21447 mus musculus
4	5301	81.9	1276	MDR1_MOUSE	P06795 mus musculus
5	5282	81.6	1276	MDR2_CRIGR	P21449 cricetus
6	5231.5	80.8	1277	MDR1_RAT	P43245 rattus norv
7	4910.5	75.8	1279	MDR3_HUMAN	P21439 homo sapien
8	4861	75.1	1276	MDR2_MOUSE	P21440 mus musculus
9	4823.5	74.5	1281	MDR3_CRIGR	P23174 cricetus
10	4808	74.3	1278	MDR2_RAT	O95342 homo sapien
11	3305	51.1	1321	AB11_HUMAN	O98201 rattus norv
12	3229.5	49.9	1321	AB11_RABIT	O980v3 oryctolagus
13	3212.5	49.6	1321	AB11_MOUSE	O9qy30 mus musculus
14	3210.5	49.6	1321	AB11_RAT	O70127 rattus norv
15	2744	42.4	1321	MDR1_CAEEL	P34712 caenorhabdi
16	2562.5	39.6	1302	MDR4_DROME	Q00449 drosophila
17	2491.5	38.5	1302	MDR5_DROME	Q00748 drosophila
18	2345	36.2	1254	MDR3_CAEEL	P34713 caenorhabdi
19	2275	35.1	1362	PM01_SCHPO	P36619 schizosacch
20	2074	32.0	1280	MDR1_LEIEN	Q06034 leishmania
21	1544	23.8	1419	MDR_PLAFF	P13568 plasmodium
22	1220.5	18.9	1290	STE6_YEAST	P12866 saccharomyc
23	1128	17.4	1336	MAM1_SCHPO	P78966 schizosacch
24	1023.5	15.8	1323	HS06_CANAL	P53706 candida alb
25	892	13.8	735	ABC8_HUMAN	Q9nut2 homo sapien
26	826	12.8	582	MSBA_ECOLI	P27299 escherichia
27	816	12.6	587	MSBA_HAEIN	P44407 haemophilus
28	772	11.9	820	MDL2_YEAST	P33311 saccharomyc
29	764	11.8	598	Y288_THEMEA	O9qytc4 thermotoga
30	761	11.8	695	MDL1_YEAST	P33310 saccharomyc
31	759	11.7	1437	MRP5_HUMAN	O15440 homo sapien
32	753.5	11.6	1436	MRP5_RAT	Q9qym0 rattus norv
33	744	11.5	1545	MRP2_HUMAN	Q92887 homo sapien

34	743.5	11.5	1436	1	MRP5_MOUSE	Q9rlx5 mus musculus
35	740.5	11.4	726	1	YFX9_SCHPO	Q9y7m7 schizosacch
36	731	11.3	1564	1	MRP2_RABIT	Q28689 oryctolagus
37	729	11.3	1325	1	MRP4_HUMAN	O15439 homo sapien
38	722	11.2	685	1	MDL1_CANAL	P97998 candida alb
39	719.5	11.1	703	1	TAP2_RAT	P36372 rattus norv
40	709	11.0	1541	1	MRP2_RAT	O63120 rattus norv
41	708	10.9	584	1	LMRA_LACLA	O9chl8 lactococcus
42	706	10.9	575	1	YNJA_BACSU	P45861 bacillus su
43	703.5	10.9	607	1	HEPA_ANASP	P22638 anabaena sp
44	699	10.8	604	1	YFIC_BACSU	P54719 bacillus su
45	696	10.8	584	1	LMRA_LACLC	P97046 lactococcus

ALIGNMENTS

RESULT 1
MDR1_HUMAN
ID MDR1_HUMAN STANDARD; PRT; 1280 AA.
AC P08183; Q12755; Q14812;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein 1).
GN ABCB1 OR PGL1 OR MDR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87028230; PubMed=2876781;
RA Chen C.-J., Chin J.E., Ueda K., Clark D.P., Pastan I., Gottesman M.M.,
Roninson I.B.;
RT "Internal duplication and homology with bacterial transport proteins
in the mdr1 (P-glycoprotein) gene from multidrug-resistant human
cells.";
RL Cell 47:381-389(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90094448; PubMed=1967175;
RA Chen C.-J., Clark D.P., Ueda K., Pastan I., Gottesman M.M.,
Roninson I.B.;
RT "Genomic organization of the human multidrug resistance (MDR1) gene
and origin of P-glycoproteins.";
RL J. Biol. Chem. 265:506-514(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97190336; PubMed=9038218;
RA Chen G., Duran G.E., Steger K.A., Lacayo N.J., Jaffrezou J.P.,
Dumontet C., Sikic B.I.;
RT "Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein,
altered phenotype, and resistance to cyclosporins.";
RL J. Biol. Chem. 272:5974-5982(1997).
RN [4]
RP SEQUENCE OF 1-234 FROM N.A.
RA Smith A., Beck C., Gibson A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 178-215 AND 800-856 FROM N.A.
RX MEDLINE=90290529; PubMed=1972623;
RA Gekeler V., Weger S., Probst H.;
RT "mdr1/P-glycoprotein gene segments analyzed from various human
leukemic cell lines exhibiting different multidrug resistance
profiles.";
RL Biochem. Biophys. Res. Commun. 169:796-802(1990).
RN [6]
RP SEQUENCE OF 1-23 FROM N.A.
RA Kioka N., Tsubota J., Kakehi Y., Komano T., Gottesman M.M.,
Pastan I., Uedak K.;
RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED

CC CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC CC -|- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, SMALL INTESTINE
CC CC AND BRAIN.
CC CC -|- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
CC CC or send an email to license@isb-slb.ch).
CC CC

DR EMBL; M14758; AAA59576.1; -
DR EMBL; M29447; AAA59576.1; -
DR EMBL; M29424; AAA59576.1; JOINED.
DR EMBL; M29425; AAA59576.1; JOINED.
DR EMBL; M29426; AAA59576.1; JOINED.
DR EMBL; M29427; AAA59576.1; JOINED.
DR EMBL; M29428; AAA59576.1; JOINED.
DR EMBL; M29429; AAA59576.1; JOINED.
DR EMBL; M29430; AAA59576.1; JOINED.
DR EMBL; M29431; AAA59576.1; JOINED.
DR EMBL; M29432; AAA59576.1; JOINED.
DR EMBL; M29433; AAA59576.1; JOINED.
DR EMBL; M29434; AAA59576.1; JOINED.
DR EMBL; M29435; AAA59576.1; JOINED.
DR EMBL; M29436; AAA59576.1; JOINED.
DR EMBL; M29437; AAA59576.1; JOINED.
DR EMBL; M29438; AAA59576.1; JOINED.
DR EMBL; M29439; AAA59576.1; JOINED.
DR EMBL; M29440; AAA59576.1; JOINED.
DR EMBL; M29441; AAA59576.1; JOINED.
DR EMBL; M29442; AAA59576.1; JOINED.
DR EMBL; M29443; AAA59576.1; JOINED.
DR EMBL; M29444; AAA59576.1; JOINED.
DR EMBL; M29445; AAA59576.1; JOINED.
DR EMBL; M29446; AAA59576.1; JOINED.
DR EMBL; AF016535; AAB69423.1; -
DR EMBL; AC002457; AAC82531.1; -
DR EMBL; M57724; AAA88047.1; -
DR EMBL; M37725; AAA88048.1; -
DR EMBL; X58723; CAA41558.1; -
DR PIR; A25059; DVH01.
DR PIR; A34914; A34914.
DR MIM; I71050; -
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmemb.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_Gtp_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 52 72 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 326 346 POTENTIAL.
FT DOMAIN 347 710 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 711 731 POTENTIAL.
FT TRANSMEM 757 777 POTENTIAL.
FT TRANSMEM 833 853 POTENTIAL.
FT TRANSMEM 854 874 POTENTIAL.
FT TRANSMEM 937 957 POTENTIAL.
FT TRANSMEM 974 994 POTENTIAL.
FT DOMAIN 995 1280 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 427 434 ATP (BY SIMILARITY).

FT	NP_BIND	1070	1077	ATP (BY SIMILARITY).
FT	REPEAT	1	637	
FT	REPEAT	638	1280	
FT	CARBOHYD	91	91	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	94	94	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	99	99	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	23	23	S -> R (IN REF. 6).
FT	CONFLICT	185	185	G -> V (IN REF. 1 AND 3).
FT	CONFLICT	336	336	MISSING (IN REF. 3).
FT	CONFLICT	412	412	G -> A (IN REF. 3).
FT	CONFLICT	438	438	O -> S (IN REF. 3).
FT	SEQUENCE	1280 AA;	141462 MW;	ABL279531F43675 CRC64;
Query Match 90.6%; Score 5866; DB 1; Length 1280;				
Best Local Similarity 90.8%; Pred. No. 0;				
Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;				
Qy	1	MDPGRGKGS	A-EKFWKGGKSKKNEKKEKPTVSTFAMFRYSNWLDRMLVGVMAAI	59
Db	1	MDLEGRNGAKKKNFKLNKSEK-DKBEKPTVSFMSFRYSNWLDRMLVGVMAAI	59	
Qy	60	IHGAAFLPLMLVFGNMTDSFANAGISRNKTFPVVINESITNTQHFINHLEEMTYAY	119	
Db	60	IHGAGLPLMLLVFGEMTDIFANAG-NLEDLMSNITRSDINDTGFPMN-LEEDMTRYAY	117	
Qy	120	YSGIGAGLVAAAYIQVSFWCLAAAGRQILKIRKOFFHAIMRQEIQWDFVDVHVGELNRLTD	179	
Db	118	YSGIGAGLVAAAYIQVSFWCLAAAGRQILKIRKOFFHAIMRQEIQWDFVDVHVGELNRLTD	177	
Qy	180	DVSKINEGIGDKGMFTQSIATFTTGIVGTRGKLTILVILAIISPVLGSLAAIAWKILS	239	
Db	178	DVSKINEGIGDKGMFTQSIATFTTGIVGTRGKLTILVILAIISPVLGSLAAIAWKILS	237	
Qy	240	SFTDKELLAYAKAGAAVEVLAAIRTVIAFGGOKKELERNKNLEAKGIGKAITANI	299	
Db	238	SFTDKELLAYAKAGAAVEVLAAIRTVIAFGGOKKELERNKNLEAKGIGKAITANI	297	
Qy	300	SIGAAFLLIYASALAFWYGTSLVLSSEYISGOVLTVPFSLVIGAFSIGOASPSIEAFAN	359	
Db	298	SIGAAFLLIYASALAFWYGTSLVLSSEYISGOVLTVPFSLVIGAFSIGOASPSIEAFAN	357	
Qy	360	ARGAAVEIFKIDNKPISIDYSKSGHKPDNKNLEPKNVHFSYPSRKEVKILKGLNKKV	419	
Db	358	ARGAAVEIFKIDNKPISIDYSKSGHKPDNKNLEPKNVHFSYPSRKEVKILKGLNKKV	417	
Qy	420	QSGOTVALVGNCGKSTTVQLMQRLYDPTDGMVCIQDQDIRTINVRLHLEITGVWSQEP	479	
Db	418	QSGOTVALVGNCGKSTTVQLMQRLYDPTDGMVCIQDQDIRTINVRLHLEITGVWSQEP	477	
Qy	480	VLFAITTAENIRYGRNVTWDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKQ	539	
Db	478	VLFAITTAENIRYGRNVTWDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKQ	537	
Qy	540	RIATARALVRNPKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADV	599	
Db	538	RIATARALVRNPKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADV	597	
Qy	600	IAGFDGIVIVEKGNHDELMKEKGIYFKLVTMTQTRGNEIELEENATGESKSDALEMSPKD	659	
Db	598	IAGFDGIVIVEKGNHDELMKEKGIYFKLVTMTQTRGNEIELEENATGESKSDALEMSPKD	657	
Qy	660	SGSLIKRRSTRRSIHAPQODRKLGTKEDLNENVPVSWRILKLNSTWPFVVGIFC	719	
Db	658	SRSLIKRRSTRRSVRSQADRKLSKALDESIPPVSWRIMKLNLTWPFVVGIFC	717	
Qy	720	AIINGGLQPAFSIIFSRIGIFTRDEDPETKRONSMFVFLVFLIGIISITFFFLQGFTE	779	
Db	718	AIINGGLQPAFSIIFSKLIGVFTRIDDPETKRONSMFVFLVFLIGIISITFFFLQGFTE	777	
Qy	780	GKAGEILTTLRLRYMVFRRSMLRQDYSWFDDPKNTTGALTTLANDAAQVKAIGSRLLAVIT	839	
Db	778	GKAGEILTTLRLRYMVFRRSMLRQDYSWFDDPKNTTGALTTLANDAAQVKAIGSRLLAVIT	837	

QY 840 QNANLGTGIIISLYGQWLTLLLLAIIVPIIAIAGVWEMKMSGQALKDKKELEGAGKIA 899
 Db 838 QNANLGTGIIISFYISYQWLTLLLLAIIVPIIAIAGVWEMKMSGQALKDKKELEGAGKIA 897
 QY 900 TEAIENFTVYSLTREQFEYMYAQSLOVPYRNSLRKAHIFGVFSFTQAMMYFSYAGCF 959
 Db 898 TEAIENFTVYSLTREQFEYMYAQSLOVPYRNSLRKAHIFGVFSFTQAMMYFSYAGCF 957
 QY 960 RFGAYLVANEFMNFVDLLVFSIAIVFGAMAVGVSSFPADYAKAKVSAHAHVIIMIEKSP 1019
 Db 958 RFGAYLVANEFMNFVDLLVFSIAIVFGAMAVGVSSFPADYAKAKVSAHAHVIIMIEKSP 1017
 QY 1020 IDSYSYHGLKNTLEGNVTFNEVFNYPTRPDIPVLOGLSLVKKGTQALVSGSGCGKS 1079
 Db 1018 IDSYSYHGLKNTLEGNVTFNEVFNYPTRPDIPVLOGLSLVKKGTQALVSGSGCGKS 1077
 QY 1080 TVQQLERFYDPLAGVLIDGKEIKHNVQWLRHGLGVSEPIFLDCSTAEINAYGDNS 1139
 Db 1078 TVQQLERFYDPLAGVLIDGKEIKHNVQWLRHGLGVSEPIFLDCSTAEINAYGDNS 1137
 QY 1140 RVVSHETMOAAKEANIHHFETIPEKYNTRVGDGTQSLGSGQKORIAIARALVQPHIL 1199
 Db 1138 RVVSHETMOAAKEANIHHFETIPEKYNTRVGDGTQSLGSGQKORIAIARALVQPHIL 1197
 QY 1200 LIDETSALDTESEKVVQVQALDKAREGTCIVIAHRLSTIQNADLIIVFQNGRYKRGTH 1259
 Db 1198 LIDETSALDTESEKVVQVQALDKAREGTCIVIAHRLSTIQNADLIIVFQNGRYKRGTH 1257
 QY 1260 QOLLAOKGIYFSMVSVQAGKR 1281
 Db 1258 QOLLAOKGIYFSMVSVQAGTKR 1279

RESULT 2
 MDL1_CRIGR STANDARD; PRT; 1276 AA.
 AC P21448;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Multidrug resistance protein 1 (P-glycoprotein 1).
 GN ABCB1 OR PGY1 OR PGPI.
 OS Cricetus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92135896; PubMed=1685679;
 RA Endicott J.A., Sarangi F., Ling V.;
 RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
 gene family.";
 RL DNA Seq. 2:89-101(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91154265; PubMed=1671863;
 RA Devine S.E., Hussain A., Davide J.P., Melera P.W.;
 RT "Full length and alternatively spliced pgp1 transcripts in multidrug-
 resistant Chinese hamster lung cells.";
 RL J. Biol. Chem. 266:4545-4555(1991).
 RN [3]
 RP SEQUENCE OF 706-1276 FROM N.A.
 RX MEDLINE=88122132; PubMed=2893255;
 RA Endicott J.A., Juranka P.F., Sarangi F., Gerlach J.H., Deuchars K.L.,
 RA Ling V.;
 RT "Simultaneous expression of two P-glycoprotein genes in
 drug-sensitive Chinese hamster ovary cells.";
 RL Mol. Cell. Biol. 7:4075-4081(1987).
 CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
 CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT

CC CAPABILITIES: PGPI AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
 CC CANNOT.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M60040; AAA68883.1; -;
 CC EMBL: M59253; AAA37004.1; -;
 CC EMBL: M17897; AAA37006.1; -;
 CC PIR: A36966; DVHYIC.
 CC InterPro: IPR003593; AAA.
 CC InterPro: IPR001140; ABC_transporter_tmem.
 CC InterPro: IPR003439; ABC_transportr.
 CC InterPro: IPR001687; ATP_GTP_A.
 CC Pfam: PF00664; ABC_membrane; 2.
 CC Pfam: PF00005; ABC_tran; 2.
 CC SMART: SM00382; AAA; 2.
 CC PROSITE: PS00211; ABC_TRANSPORTER; 2.
 CC KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
 KW Multigene family.
 FT DOMAIN 1 50 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 213 233 POTENTIAL.
 FT TRANSMEM 294 314 POTENTIAL.
 FT TRANSMEM 323 343 POTENTIAL.
 FT TRANSMEM 344 707 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 708 774 POTENTIAL.
 FT TRANSMEM 754 774 POTENTIAL.
 FT TRANSMEM 830 850 POTENTIAL.
 FT TRANSMEM 851 871 POTENTIAL.
 FT TRANSMEM 934 954 POTENTIAL.
 FT TRANSMEM 971 991 POTENTIAL.
 FT DOMAIN 992 1276 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 424 431 ATP (POTENTIAL).
 FT NP_BIND 1067 1074 ATP (POTENTIAL).
 FT REPEAT 1 635
 FT REPEAT 636 1276
 FT CONFLICT 338 339
 FT SEQUENCE 1276 AA; 140925 MW; 44F3F92A186B4DFF CRC64;
 SQ
 Query Match 87.5%; Score 5667.5; DB 1; Length 1276;
 Best Local Similarity 87.1%; Pred. No. 4.5e-316;
 Matches 1116; Conservative 77; Mismatches 82; Indels 7; Gaps 3;
 QY 1 MDPGGKRSKAEKNFWKMGKSKKNEKPKPTVSTFAMFRYSNNMLDRLMLVGTMAAI 60
 Db 1 MEFEDEFSGRKDKNFKMGKSKK-EKKEKPKVSVFTMFYAGWLDRLMLVGTMAAI 59
 QY 61 HGAALPLMLVFGNMTDSPANAG-ISRKNKTPFVPIINESITNTQHFHINLEEMTYAY 119
 Db 60 HGVALLPLMLVFGNMTDSPANAG-ISRKNKTPFVPIINESITNTQHFHINLEEMTYAY 114
 QY 120 YSGIGAGVLVAAYIOVSFWCLAAAGQILKIRKOFFHAIMRQEIQWFDVHDVGLNRLTD 179
 Db 115 YTGIGAGVLIVAYIOVSFWCLAAAGQILKIRKOFFHAIMRQEIQWFDVHDVGLNRLTD 174
 QY 180 DYSKINEGIDKIGMFFQSIAITFTGIVGFTGRGKLTIVILAIISPVLSAGIAWAKILS 239
 Db 175 DYSKINEGIDKIGMFFQSIAITFTGIVGFTGRGKLTIVILAIISPVLSAGIAWAKILS 234
 QY 240 SFTDELLAYAKAGAAVEVLAAIRTVTAFGQKKELERYKNLEAKGIGIKKAITANI 299
 Db 235 SFTDELLAYAKAGAAVEVLAAIRTVTAFGQKKELERYKNLEAKGIGIKKAITANI 294
 QY 300 SIGAAFLLIYASYALAFWYGTSLVLSSEYSIGQVLTVFVSFLIGAFSGSPSEAFAN 359

Db	295	SMGAFLLIYASYALAFWGTSLVSKYSIGOVLTVEFAVLIGAFSGQASPNIEAFAN	354	DT	01-APR-1993 (Rel. 25, Last sequence update)
Qy	360	ARGAAEYFKIIDNKPSIDSKYKSHKPDNTKGNLFKNVHFSPSRKEVKILKGLNLKV	419	DE	16-OCT-2001 (Rel. 40, Last annotation update)
Db	355	ARGAAEYFNIIDNKPSIDSKYKSHKPDNTKGNLFKNVHFSPSRKDVQLKGLNLKV	414	GN	ABCBA OR PGY3 OR PGY-3 OR MDR3 OR MDR1A.
Qy	420	QSGQVALVGNCGCKSTTVQLMORLYDPTDGMVCDIGQDRTINVRHLREITGVVSGEP	479	OC	Mus musculus (Mouse).
Db	415	QSGQVALVGNCGCKSTTVQLMORLYDPTDGMVCDIGQDRTINVRHLREITGVVSGEP	474	OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Qy	480	VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDFLGVGERGAQLSGGQKQ	539	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Db	475	VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDFLGVGERGAQLSGGQKQ	534	OX	NCBI_TaxID=10090;
Qy	540	RTAIALRVNPKILLDEATSALETSADVAVQVADLKARKGRTTIVIAHRLSTVRNADV	599	RP	SEQUENCE FROM N.A.
Db	535	RTAIALRVNPKILLDEATSALETSADVAVQVADLKARKGRTTIVIAHRLSTVRNADI	594	RX	MEDLINE=90205845; PubMed=1969610;
Qy	600	IAGFDDGVIVKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKD	659	RA	Devault A., Gros P.;
Db	595	IAFGDGVIVKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKD	654	RA	"Two members of the mouse mdr gene family confer multidrug resistance
Qy	660	SGSSLKRSTRSRSHAPQDQKLGTKEDLNENVPVPSFWIRILKLNSTWPFYFVGIFC	719	RT	with overlapping but distinct drug specificities.";
Db	655	SASSLLRRSTRSRSHAPQDQKLGTKEDLNENVPVPSFWIRILKLNSTWPFYFVGIFC	714	RT	Mol. Cell. Biol. 10:1652-1663(1990).
Qy	720	AIINGLOPAFSTIESRIIGTIFREDPETKRONSMFSLVFLVGLIISFTFFLQGTFF	779	RN	[2]
Db	715	AIINGLOPAFSTIESRIIGTIFREDPETKRONSMFSLVFLVGLIISFTFFLQGTFF	774	RP	SEQUENCE FROM N.A.
Qy	780	GRAGEILTRRLRYMFRSMRDQVSWDFDPPKNTTGALTTRLANDAAQVKGATGSLAVIT	839	RX	MEDLINE=90287150; PubMed=1972547;
Db	775	GRAGEILTRRLRYMFRSMRDQVSWDFDPPKNTTGALTTRLANDAAQVKGATGSLAVIT	834	RA	Hsu S.I.H., Cohen D., Kirschner L.S., Lothstein L., Hartstein M.,
Qy	840	QNTANLGTGIISLYGWOLTLALLAIVPIIAIAGVVENKMLSGQALDKKELEGAGKIA	899	RA	Horwitz S.B.;
Db	835	QNTANLGTGIISLYGWOLTLALLAIVPIIAIAGVVENKMLSGQALDKKELEGAGKIA	894	RT	"Structural analysis of the mouse mdr1a (P-glycoprotein) promoter
Qy	900	TEAENFRVWSLTROKFEYVYQSLVPRYNSLRKKAHFGVSTFSTQAMMYFAGCF	959	RT	reveals the basis for differential transcript heterogeneity in
Db	895	TEAENFRVWSLTROKFEYVYQSLVPRYNSLRKKAHFGVSTFSTQAMMYFAGCF	954	RT	multidrug-resistant J774.2 cells.";
Qy	960	RFGAYLVANEFMFDQVLLVFSALVFGAMAVGVSSFPADYAKAVSAHVIMIEKSP	1019	RT	Mol. Cell. Biol. 10:3596-3606(1990).
Db	955	RFGAYLVANEFMFDQVLLVFSALVFGAMAVGVSSFPADYAKAVSAHVIMIEKSP	1014	RN	[3]
Qy	1020	IDSYSPLKPNTEGNTFNEVFNYPTRPDIPVQLGSLVKKQTLALVGS	1079	RP	SEQUENCE OF 173-1276 FROM N.A.
Db	1015	IDSYSPLKPNTEGNTFNEVFNYPTRPDIPVQLGSLVKKQTLALVGS	1074	RC	STRAIN=BALB/C;
Qy	1080	TVVOLLERFVPLAGSVLDGKEIKHLNQLVWLRHGLVSVQPIILFDCSIAENIAYGNS	1139	RX	MEDLINE=89308614; PubMed=2473069;
Db	1075	TVVOLLERFVPLAGSVLDGKEIKHLNQLVWLRHGLVSVQPIILFDCSIAENIAYGNS	1134	RA	Hsu S.I.H., Lothstein L., Horwitz S.B.;
Qy	1140	RVVSHEEIMQAAKEANIHFIETLPKYNTRVGDGKTQLSGGQKQRIATARALVRPHIL	1199	RA	"Differential overexpression of three mdr gene family members in
Db	1135	RVVSQDEIERAAKEANIHQIESLPDKYNTRVGDGKTQLSGGQKQRIATARALVRPHIL	1194	RT	multidrug-resistant J774.2 mouse cells. Evidence that distinct P-
Qy	1200	LLDEATSALETSADVAVQVADLKARKGRTTIVIAHRLSTVRNADVAVQVADLKARKGRTTIVIAHRLSTVRNADV	1259	RT	glycoprotein precursors are encoded by unique mdr genes.";
Db	1195	LLDEATSALETSADVAVQVADLKARKGRTTIVIAHRLSTVRNADVAVQVADLKARKGRTTIVIAHRLSTVRNADI	1254	RL	J. Biol. Chem. 264:12053-12062(1989).
Qy	1260	QQLAOKGIYFMSVQAGAKR 1281		CC	-1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
Db	1255	QQLAOKGIYFMSVQAGAKR 1276		CC	DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
RESULT 3				CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.
MDR3_MOUSE				CC	-1- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
ID MDR3_MOUSE				CC	RELATED BUT DISTINCT CELLULAR GENES.
AC P21447;				CC	-1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
DT 01-MAY-1991 (Rel. 18, Created)				CC	-----
				CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
				CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
				CC	the European Bioinformatics Institute. There are no restrictions on its
				CC	use by non-profit institutions as long as its content is in no way
				CC	modified and this statement is not removed. Usage by and for commercial
				CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
				CC	or send an email to license@isb-sib.ch).
				CC	-----
				CC	EMBL; M30697; AAA39517.1; -
				DR	EMBL; M33581; AAA39514.1; -
				DR	EMBL; M33580; AAA39518.1; -
				DR	EMBL; M24417; AAA03243.1; -
				DR	PIR; A34175; DVMS1A.
				DR	PIR; A34786; A34786.
				DR	MGD; MGI:97570; Abcb4
				DR	InterPro; IPR003593; AAA.
				DR	InterPro; IPR001140; ABC_transporter_tmnm.
				DR	InterPro; IPR003439; ABC_transporter.
				DR	InterPro; IPR001687; ATP-GTP_A.
				DR	Pfam; PF00664; ABC_membrane; 2.
				DR	Pfam; PF00005; ABC_tran; 2.
				DR	SMART; SM00382; AAA; 2.
				DR	PROSITE; PS00211; ABC_TRANSPORTER; 2.
				DR	ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
				KW	Multigene family.
				FT	DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
				FT	TRANSMEM 71 POTENTIAL.
				FT	TRANSMEM 117 136 POTENTIAL.
				FT	TRANSMEM 188 205 POTENTIAL.
				FT	TRANSMEM 212 232 POTENTIAL.
				FT	TRANSMEM 298 318 POTENTIAL.
				FT	TRANSMEM 327 346 POTENTIAL.
				FT	DOMAIN 347 707 CYTOPLASMIC (POTENTIAL).

```

FT TRANSMEM 708 728 POTENTIAL.
FT TRANSMEM 753 773 POTENTIAL.
FT TRANSMEM 829 848 POTENTIAL.
FT TRANSMEM 853 880 POTENTIAL.
FT TRANSMEM 942 961 POTENTIAL.
FT TRANSMEM 964 984 POTENTIAL.
FT DOMAIN 985 1276 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 423 430 ATP (BY SIMILARITY).
FT NP_BIND 1066 1073 ATP (BY SIMILARITY).
FT REPEAT 1 632
FT REPEAT 633 1276
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 526 527 QL -> HV (IN REF. 2 AND 3).
FT CONFLICT 939 939 F -> S (IN REF. 2 AND 3).
FT CONFLICT 1036 1036 F -> V (IN REF. 2 AND 3).
SQ SEQUENCE 1276 AA; 140754 MW; 75C71F33E1F58481 CRC64;

Query Match 87.1%; Score 5638; DB 1; Length 1276;
Best Local Similarity 87.1%; Pred. No. 2.2e-314;
Matches 1116; Conservative 74; Mismatches 84; Indels 8; Gaps 4;

Qy 1 MDPEGRKGSAEKFWKMGKSKKNEKKKPTVSTFAMFRYSNWLDRMLVGLTMAAII 60
Db 1 MELEEDLKGRADKNFKSMGKSKK- EKKEKPAVSVLTMFRYAGWLDRLYMLVGLTAAII 59

Qy 61 HGAALPLMLVFGNWTGSFANAG-ISRKNTPPVLINESITNNTQHFHNHLEEMTTVAYY 119
Db 60 HGVALPLMLVFGDMTDSFASGVNKNST-----NNSEADKRAMFAK-LEEEMTTVAYY 113

Qy 120 YSGIGAGVLVAAIYQVSWFCLAAQRILKIRKQFFHAIMRQEIQWFDVHDVDELNRLTD 179
Db 114 YTGIGAGVLVAAIYQVSWFCLAAQRILKIRKQFFHAIMRQEIQWFDVHDVDELNRLTD 173

Qy 180 DYSKINGIGDKIGMFTQSTATFTTGTGIVGTRGKWLTVLILASPVGLISAAIWAIIKS 239
Db 174 DYSKINGIGDKIGMFTQAMATFTGFTIGTRGKWLTVLILASPVGLISAGIWAIIKS 233

Qy 240 SFTDKELLAYAKAGAAVEVLAAIRTVIAFGQKKELERYKNLEEAKGIGIKKAITANI 299
Db 234 SFTDKELHAYAKAGAAVEVLAAIRTVIAFGQKKELERYNNLEEAKRIGIKKAITANI 293

Qy 300 SIGAAFLIYASYALAFWYGTSLVLSSEYSIGQVLTIVFFSVLIGAFSIGQASPSIEAFAN 359
Db 294 SMGAFLIYASYALAFWYGTSLVLSSEYSIGQVLTIVFFSVLIGAFSVGASPSIEAFAN 353

Qy 360 ARGAAEIFIKIDNKPIDSYSKSGHPDNKGNLEPKNVHFSYPSRKEVKILKGLNLKV 419
Db 354 ARGAAEYFVFIIDNKPIDSYSKSGHPDNKGNLEPKNVHFSYPSRKEVKILKGLNLKV 413

Qy 420 QSGQTVLVGNSGCGKSTTVQLMORLYDPTDGMVCIDGQDIRTINVRHLREITGVVSOEP 479
Db 414 KSGQTVLVGNSGCGKSTTVQLMORLYDPTDGMVSDIGQDIRTINVRHLREITGVVSOEP 473

Qy 480 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKEDTLVGERGAOLSGGQKQ 539
Db 474 VLFATTIAENIRYGREVDVTDEIEKAVKEANAYDFIMKLPNQFDTLVGERGAOLSGGQKQ 533

Qy 540 RIATARALVRNPKILLDDEATSDALDESEAVQVQALDKARKGRTTIVIAHRLSTVRNADV 599
Db 534 RIATARALVRNPKILLDDEATSDALDESEAVQVQALDKAREGRTTIVIAHRLSTVRNADV 593

Qy 600 IAGFDGCVIYKGNHDELMKEGIFKLVMTQTRGNEIELENATGSKESDESDALEMSPKD 659
Db 594 IAGFDGCVIYKGNHDELMKEGIFKLVMTQTAGNEIELGNACKSKDEIDMLDSSKD 653

Qy 660 SSSLIKRRSTRRSIHAPOQDRKLTGEDLNENVPVPSFWRILKLNSTWPFYVVGIFC 719
Db 654 SSSLIKRRSTRRSIHAPOQDRKLTGEDLNENVPVPSFWRILKLNSTWPFYVVGIFC 713

Qy 720 AILINGGLQPAFSAIFSRIGIFTRDEDEPKTRQNSNMFVLFVLVLVLIIGISFIITFFLQGTFF 779
Db 714 AILINGGLQPAFSAIFSKVGVETNGGPPETQRONSNLFSLFLILIGISFIITFFLQGTFF 773
Db 780 GKAGEILTKRLRYWVFRSMRLQDVSWFDDPKNTTGALTITRLANDAQAQVGAIGSRVAVIT 839
Db 774 GKAGEILTKRLRYWVFRSMRLQDVSWFDDPKNTTGALTITRLANDAQAQVGAIGSRVAVIT 833
Qy 840 QNTANLGTGIIISLIYIGWQTLTLLLAIVPIIAIAGVVMKMLSGQALKKKKELEGAGKIA 899
Db 834 QNTANLGTGIIISLIYIGWQTLTLLLAIVPIIAIAGVVMKMLSGQALKKKKELEGAGKIA 893
Qy 900 TEAIEFNRTVSLTREOKFEYMYAQSLQVPRYNSLRKAHIFGVYSFISITQAMVFSYAGCF 959
Db 894 TEAIEFNRTVSLTREOKFETMYAQSLQVPRYNSLRKAHIFGVITFTTQAMVFSYAACF 953
Qy 960 RFAGAYLVANEFMFQDVLVLFSAIVFAGAMAVGVSSFADPDYAKAKVSAAHVIMIEKSPIL 1019
Db 954 RFAGAYLVTOQLMTFFENLVLFSAIVFAGAMAVGVSSFADPDYAKATVSASHIIRIEKTPPE 1013
Qy 1020 IDSYSPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSSSCGCKS 1079
Db 1014 IDSYSTQGLKPNMNEGVOFSGFVFNYPTRPSIPVLQGLSLEVKKGOTLALVSSSCGCKS 1073
Qy 1080 TVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHILGIVSQEPIFDFCSIAENIAYGDNS 1139
Db 1074 TVVOLLERFYDPMWAGSVLIDGKEIKHLNVQWLRHILGIVSQEPIFDFCSIAENIAYGDNS 1133
Qy 1140 RVYSHEIEIQAQAEKANIHHIFETLPEKYNTRYVGDGKTQLSGGOKQRIATARALVRQPHIL 1199
Db 1134 RVYSHEIEIQAQAEKANIHHIFIDSLDPKYNTRYVGDGKTQLSGGOKQRIATARALVRQPHIL 1193
Qy 1200 LLDEATSAIDTESEKVVQVQALDKAREGRTTIVIAHRLSTIQNALDLIVVQNGKVEHGTH 1259
Db 1194 LLDEATSAIDTESEKVVQVQALDKAREGRTTIVIAHRLSTIQNALDLIVVQNGKVEHGTH 1253
Qy 1260 QQLLAQKQIYFMSVSVQAGAKR 1281
Db 1254 QQLLAQKQIYFMSVSVQAGAKR 1275

RESULT 4
MDRL_MOUSE STANDARD; PRT; 1276 AA.
ID MDRL_MOUSE AC P06795;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein 1).
GN ABCB1 OR PGY1 OR PGY1-1 OR MDR1 OR MDR1B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87028229; PubMed=3768958;
RA Gros P., Croop J., Housman D.;
RT "Mammalian multidrug resistance gene: complete cDNA sequence
indicates strong homology to bacterial transport proteins.";
RL Cell 47:371-380(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89367274; PubMed=2570420;
RA Raymond M., Gros P.;
RT "Mammalian multidrug-resistance gene: correlation of exon
organization with structural domains and duplication of an ancestral
gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:6488-6492(1989).
RN [3]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=91042535; PubMed=2248681;
RA Raymond M., Gros P.;
RT "Cell-specific activity of cis-acting regulatory elements in the
promoter of the mouse multidrug resistance gene mdr1.";

```


QY 901 EAIENFRTVSLTRQKFEYMAQSLQVYPYRSLRAKHIFGVFSFISQAMMYFSYAGCFR 960
DB 897 EAIENFRTVSLTRQKFEYMAQSLQVYPYRSLRAKHIFGVFSFISQAMMYFSYAGCFR 956
QY 961 FGAYLVANEFMFQDVLVFSYAIVFGAMAVGVSSFADPDYAKAKVSAHVIMITEKSLPI 1020
DB 957 FGAYLVHQIMTFENVMVFSVAVFGAAGNASSFADPDYAKAKVSAHVIMITEKSLPI 1016
QY 1021 DSYSPHGLKPNVLEGNVFNVEVNPTRPDIPVLQGLSLEVKQGTALVGVSSCGKST 1080
DB 1017 DSYSPHGLKPNVLEGNVFNVEVNPTRPDIPVLQGLSLEVKQGTALVGVSSCGKST 1076
QY 1081 VVQLLEFYDPLAGSVLIDGKEIKHLNVQWLAHLGIVSQEPILFDCSIAENIAYGDNRS 1140
DB 1077 VVQLLEFYDPMAGTVLDGKEIKHLNVQWLAHLGIVSQEPILFDCSIAENIAYGDNRS 1136
QY 1141 VVSHEIMQAQKANIHHFETLPKYNTRVGDGKTQLSGGQKQRIATARALVROPHILL 1200
DB 1137 VVSQDEIERAAKEANTHPIESLPKYNTRVGDGKTQLSGGQKQRIATARALVROPHILL 1196
QY 1201 LDENATSDTESKVVQVQALDKARGRTCVIAHRLSTIONADLVVFNQGVKKEHGHQ 1260
DB 1197 LDENATSDTESKVVQVQALDKARGRTCVIAHRLSTIONADLVVFNQGVKKEHGHQ 1256
QY 1261 QLLAQKGIYFSMVQAGAKR 1281
DB 1257 QLLAQKGIYFSM--VQAGAKR 1275
RESULT 6
MDR1.RAT STANDARD; PRT: 1277 AA.
AC P43245.
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein 1).
GN ABCB1 OR P-gly OR MDR1 OR MDR1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92039081; PubMed=1682220;
RA Silverman J A., Raunio H., Gant T.W., Thorgeirsson S.S.;
RT "Cloning and characterization of a member of the rat multidrug
RT resistance (mdr) gene family.";
RL Gene 106:229-236(1991).
CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- PTM: SEVERAL PHOSPHORYLATED SERINE RESIDUES ARE PRESENT IN THE
CC LINKER DOMAIN.
CC -1- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
CC RELATED BUT DISTINCT CELLULAR GENES.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81855; -; NOT_ANNOTATED_CDS.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC transporter_tmern.
DR InterPro: IPR003439; ABC transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane; 2.
DR Pfam: PF00005; ABC_tran; 2.
DR

SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 48 71 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 327 347 POTENTIAL.
FT TRANSMEM 348 709 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 710 730 POTENTIAL.
FT TRANSMEM 755 775 POTENTIAL.
FT TRANSMEM 832 852 POTENTIAL.
FT TRANSMEM 854 874 POTENTIAL.
FT TRANSMEM 937 957 POTENTIAL.
FT TRANSMEM 968 988 POTENTIAL.
FT DOMAIN 989 1277 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (BY SIMILARITY).
FT NP_BIND 1070 1077 ATP (POTENTIAL).
FT REPEAT 1 635
FT REPEAT 636 1277
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1277 AA; 141386 MW; 8AFDDDD619D2934C1 CRC64;
Query Match 80.8%; Score 5231.5; DB 1; Length 1277;
Best Local Similarity 80.0%; Pred. No. 3.6e-291;
Matches 1029; Conservative 131; Mismatches 111; Indels 15; Gaps 9;
QY 1 MDEGGRKGSAAKFNWKMGGKSKKNEKKKPTVSTFAMPRYSNWLDRMLVMTGTHAAII 60
DB 1 MEFEGLNGRADKFNFSKMGKSKK--EKERKPAVGFEGMFYADWLDKCLMGALGTAAII 58
QY 61 HGAALPLMLVFGNMTDSFANAGISRN-KTFPPVIINESITNTOHETN--HLSEEMTYA 117
DB 59 HGTLLPLMLVFGNMTDSF--SRDPHSORAITNTOSEINST-HIVSDTSLEEDMAYIA 114
QY 118 YYSIGGAGVLAAYIQVSEFWCLAAAGRIKIRKQFFHAIMRQIEIGWFDVHDVDELNRL 177
DB 115 YYYTGIGAGVLAAYIQVSLWCLAAAGRIKIRKQFFHAIMRQIEIGWFDVNDAGELNRL 174
QY 178 TDDVSKNEGIGDKIMFFQFSIATFFTFVGTFRGKLTIVLAISPVLGLSAAIWKI 237
DB 175 TDDVSKNEGIGDKIMFFQFSIATFFTFVGTFRGKLTIVLAISPVLGLSAAIWKI 234
QY 238 LSSFTDKELLAYAKAGAAVEVLAAITVIAFGGKKELERYKNKLEAKGIGIKKAITA 297
DB 235 LTFNKLQAYAKAGAAVEVLAAITVIAFGGKKELERYKNKLEAKGIGIKKAITA 294
QY 298 NISGAAPLLIYAYALAFWVGTSLVLSSEYSIGQVLTVFFSVLIGAFSGQASPSPEAF 357
DB 295 NISGIAVLLVYAYALAFWVGTSLVLSSEYSIGQVLTVFFSVLIGAFSGQASPSPEAF 354
QY 358 ANAGAAYEIKIIDNKPISIDYSKSGHKPDNIKNGLEFNKVFHSPSRKVKILKGLN 417
DB 355 ANAGAAYEIKIIDNKPISIDYSKSGHKPDNIKNGLEFNKVFHSPSRKVKILKGLN 414
QY 418 KVQSGQTVALVGNCGCKSTTVQLMQRILYDPTDGMVCDIGQDITINVRHLREITGVVSQ 477
DB 415 KVQSGQTVALVGNCGCKSTTVQLMQRILYDPTDGMVCDIGQDITINVRHLREITGVVSQ 474
QY 478 EPLVFAITIAENIRYGRNVMTDEIEKAVKANAYDFIMKLPNKFDFLVERGAQLSGGQ 537
DB 475 EPLVFAITIAENIRYGRNVMTDEIEKAVKANAYDFIMKLPNKFDFLVERGAQLSGGQ 534
QY 538 KORTAIALVRNPKILLDEATSAIDTESAVVQVQALDKARGRTTVIAHRLSTVYRNA 597
DB 535 KORTAIALVRNPKILLDEATSAIDTESAVVQVQALDKARGRTTVIAHRLSTVYRNA 594
QY 598 DVIAGFDGVIIVEKGNHDELMEKGIYFKLVMTQTRNGNEIELENATGESADLEMS 657
DB 595 DVIAGFDGVIIVEKGNHDELMEKGIYFKLVMTQTRNGNEIELENATGESADLEMS 654

```
QY 658 KDSSSLIKRRSTRSHAPQDQKRLKEDLNENVPVSWFRLKLNSTWEPYFVVG 717
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 655 EESKPLI-RSIRSIHRQDQERRLSKEDVEDPMVSWFQWILKLNISEWYLVGV 713
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 718 FCAINGLOPAPSIIFRSIIGITRDEPPTKQNSMFMVSLFLVLGIISFIFFLQGF 777
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 714 LCAVINGCIQVFAIVFSKIVGVFSRDDHETKQRCNLCNLFSLFLVMGMISFVYFFQGF 773
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 778 TFGKAGEILTRLRYMVPFRSMLRDVSNFDDPKNTTGATLTRLANDAAQVKGATGSLAV 837
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 774 TFGKAGEILTRLRYMVPFRSMLRDVSNFDDPKNTTGATLTRLANDAAQVKGATGSLAV 833
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 838 ITQNIANTGIIIS--LIYQWLTLLLAIVP:IAIAGVVMKLSQALKDKKELEGA 895
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 834 VTQVANLGTGIIISLVLYGQWLTLLVWIPIVLVGGIIEMLLSQALKDKKELEIS 893
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 896 GKIAEALERTVVSILTREQKFYMVAQSLQVYPRNSLRKAHIFGVFSFISITQAMTFYS 955
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 894 GKIAEALERTVVSILTREQKFYMVAQSLQVYPRNSLRKAHIFGVFSFISITQAMTFYS 953
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 956 ACGRFGAYLVANFMFQDVLVFSALVFGAMAVGOVSSFPDYAKAKYSAAHVIMIE 1015
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 954 AACRFGAYLVARELMTFENVMVFSVAVFGAMAAGNTSSFPDYAKAKVSASHIIGIE 1013
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1016 KSPIDSYSPHGLKNTLEGNVTNEVFNYPTRPDIPVLQGLSLEVYKGGTTLALVGSSG 1075
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1014 KIPEDSYSTBGLKPNWLEGNVKNYKFNYPTRPNIPVLQGLSFEVYKGGTTLALVGSSG 1073
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1076 CGKSTVQLLRFYDPLAGSLVDIGKEIKHLNWOLRAHAGTIVSQEPIILFDCSAENIAY 1135
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1074 CGKSTVQLLRFYDPLAGSLVDIGKEIKHLNWOLRAHAGTIVSQEPIILFDCSAENIAY 1132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1136 GDSRVVSHEEIMQAKEANIHFIETLPEKYNTRVGDKGTOLSGGQKQRIATARALVRQ 1195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1133 GDSRVVSHEEIVRAEAREANHIFDLSPEKYNTRVGDKGTOLSGGQKQRIATARALVRQ 1192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1196 PHILLDEATSALDSEKVVQEAALDKAREGTCIVIAHRLSTIQNADLVVVFQNGKVKE 1255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1193 PHILLDEATSALDSEKVVQEAALDKAREGTCIVIAHRLSTIQNADLVVVFQNGKVKE 1252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1256 HGTHQQLLAQKGIYFSMVSVQAGAKR 1281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1253 HGTHQQLLAQKGIYFSM--VOAGAKR 1276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 7
MDR3_HUMAN STANDARD; PRT; 1279 AA.
AC P21439;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (P-glycoprotein 3).
GN ABCB4 OR PGY3 OR MDR3.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89138016; PubMed=2906314;
RA van der Bliek A.M., Koolman P.M., Schneider C., Borst P.;
RT "Sequence of mdr3 cDNA encoding a human P-glycoprotein.";
RL Gene 71:401-411(1988).
RN [2]
RP SEQUENCE OF 856-1279 FROM N.A.
RX MEDLINE=88111519; PubMed=2892668;
RA van der Bliek A.M., Baas F., ten Houte de Lange T., Koolman P.M.,
RA van der Velde-Koerts T., Borst P.;
RT "The human mdr3 gene encodes a novel P-glycoprotein homologue and
RT gives rise to alternatively spliced mRNAs in liver.";
RL EMBO J. 6:3325-3331(1987).
```

```
RN [3]
RP GENE STRUCTURE.
RX MEDLINE=91161629; PubMed=2002063;
RA Lincke C.R., Smit J.J.M., van der Velde-Koerts T., Borst P.;
RT "Structure of the human MDR3 gene and physical mapping of the human
RT MDR locus.";
RL J. Biol. Chem. 266:5303-5310(1991).
CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS. HUMAN MDR3 IS NOT
CC CAPABLE OF CONFERRING DRUG RESISTANCE. MEDIATES THE TRANSLLOCATION
CC OF PHOSPHATIDYLCHOLINE ACROSS THE CANALICULAR MEMBRANE OF THE
CC HEPATOCYTE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- DISEASE: DEFECTS IN ABCB4 ARE A CAUSE OF PROGRESSIVE FAMILIAL
CC INTRAHEPATIC CHOLESTASIS TYPE III (PFIC), A FORM OF AUTOSOMAL
CC RECESSIVE LIVER DISORDERS, CHARACTERIZED BY EARLY ONSET OF
CC CHOLESTASIS THAT PROGRESSES TO CIRRHOSIS AND LIVER FAILURE BEFORE
CC ADULTHOOD.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M23234; AAA36207.1; -.
DR EMBL; X06181; CAA29547.1; -.
DR PIR; JS0051; DVH03.
DR PIR; A42213; A42213.
DR HSP; P13569; LNB0.
DR MIM; 171060; -.
DR MIM; 602347; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC transporter_tmem.
DR InterPro; IPR003439; ABC transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00564; ABC_membrane; 2.
DR Pfam; SM00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 192 211 POTENTIAL.
FT TRANSMEM 216 235 POTENTIAL.
FT TRANSMEM 301 320 POTENTIAL.
FT TRANSMEM 336 354 POTENTIAL.
FT DOMAIN 355 711 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 712 732 POTENTIAL.
FT TRANSMEM 756 776 POTENTIAL.
FT TRANSMEM 832 851 POTENTIAL.
FT TRANSMEM 854 873 POTENTIAL.
FT TRANSMEM 937 956 POTENTIAL.
FT TRANSMEM 976 993 POTENTIAL.
FT DOMAIN 994 1279 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 429 436 ATP (BY SIMILARITY).
FT NP_BIND 1069 1076 ATP (BY SIMILARITY).
FT REPEAT 1 640
FT REPEAT 641 1279
FT CARBOHYD 91 91
FT CARBOHYD 97 97
FT CONFLICT 1093 1093
SQ SEQUENCE 1279 AA; 140682 MW; 3D58C98B5C8D6087 CRC64;
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
V -> VFVDFGFQ (IN REF. 2).
```

Query Match 75.8%; Score 4910.5; DB 1; Length 1279;
Best Local Similarity 75.6%; Pred. No. 8e-273;
Matches 974; Conservative 131; Mismatches 164; Indels 19; Gaps 9;

FT	TRANSMEM	52	75	POTENTIAL.
FT	TRANSMEM	119	139	POTENTIAL.
FT	TRANSMEM	188	208	POTENTIAL.
FT	TRANSMEM	215	235	POTENTIAL.
FT	TRANSMEM	296	316	POTENTIAL.
FT	TRANSMEM	325	345	POTENTIAL.
FT	DOMAIN	346	707	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	708	728	POTENTIAL.
FT	TRANSMEM	753	773	POTENTIAL.
FT	TRANSMEM	829	849	POTENTIAL.
FT	TRANSMEM	850	870	POTENTIAL.
FT	TRANSMEM	933	953	POTENTIAL.
FT	TRANSMEM	970	990	POTENTIAL.
FT	DOMAIN	991	1276	CYTOPLASMIC (POTENTIAL).
FT	NP_BIND	426	433	ATP (BY SIMILARITY).
FT	NP_BIND	1066	1073	ATP (BY SIMILARITY).
FT	REPEAT	1	633	
FT	REPEAT	636	1276	
FT	SEQUENCE	1276	AA; 140332 MW; A6C38DCD2C118EEF CRC64;	
Query Match 75.1%; Score 4861; DB 1; Length 1276;				
Best Local Similarity 74.8%; Pred. No. 5.4e-270;				
Matches 960; Conservative 139; Mismatches 171; Indels 14; Gaps 8;				
QY	1	MDPEGRKGSAAK--NFWKMKKSKK--NEKKKKPTVSTFAMPRYSNWLDRMLVGTWA	57	
DB	1	MDLEAARNGTARRLDGDFELGSIISNQREKKKVNLIIGLLTLFRYSQWQDKLFMFLGTL	60	
QY	58	AIHGAALPLMLVFGNWDSPA--NAGISRNKTFPPVINESINNTQHFINHLEEMTY	116	
DB	61	AIHAGSLPLMLVFGEMTDKFDYDNTG---NFSPLPNFSLMLNPGK---ILEEMTRY	113	
QY	117	AYYSIGAGVLAAYVQVSWFCLAAAGROILKIRKOFHAIHQEOTGWDVHDVGLNTR	176	
DB	114	AYYSGUGGVLAAYVQVSWFCLAAAGROILKIRKOFHAIHQEOTGWDVHDVGLNTR	173	
QY	177	LTDVSKINBGIGDKIMFQSIATFTFTGIVGTRGKLTILVLAISPVGLSAAIWA	236	
DB	174	LTDVSKISBGIGDKVGMFQSIATFTFTGIVGTRGKLTILVLAISPVGLSAAIWA	233	
QY	237	ILSSFTDKELLAYAKAGAAVEEVLAIIRVIAFGGQKELERYNKNLEBAKIGIKAT	296	
DB	234	ILSTFSDKELAYAKAGAAVEEVLAIIRVIAFGGQKELERYNKNLEBAKIGIKAT	293	
QY	297	ANISIGAAFLIYASVALAFWGTSLVLSSEYSIGQVLFVSVLIGAFSIOQASPIEA	356	
DB	294	ANISMGIAFLIYASVALAFWGTSLVLSSEYSIGQVLFVSVLIGAFSIOQASPIEA	353	
QY	357	FANARGAAYEIKIIDNKPISIDYSKSHGPNKIGNLEFNHVSYPKSKVKILKGLN	416	
DB	354	FANARGAAYEIKIIDNKPISIDYSKSHGPNKIGNLEFNHVSYPKSKVKILKGLN	413	
QY	417	LKVSQQTVALVGNSSCGSTTVQLMQRLYDPTDGMVCIDGQDIRTINVRHLREITGVVS	476	
DB	414	LKVSQQTVALVGNSSCGSTTVQLMQRLYDPTDGMVCIDGQDIRTINVRHLREITGVVS	473	
QY	477	QEPVLFTATTAENIRYGRNVMTDETEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSG	536	
DB	474	QEPVLFTATTAENIRYGRNVMTDETEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSG	533	
QY	537	QKORIAARALVRNPKILLDEATSDALDTESEAVVQVALDKARGRTTIVIAHRLSTVRN	596	
DB	534	QKORIAARALVRNPKILLDEATSDALDTESEAVVQVALDKARGRTTIVIAHRLSTVRN	593	
QY	597	ADVIAGFDGCVIVEKGNHDELKKEGIYFKLVMTQTRGNEIELEENATGESDALEMS	656	
DB	594	ADVIAGFDGCVIVEKGNHDELKKEGIYFKLVMTQTRGNEIELEENATGESDALEMS	653	
QY	657	PKDSGSLKRRSTRSIIHAPQOGQDKLGTEDLNENVPYFVFWILKLNSTEWFFVVG	716	
DB	654	P-NGWKARFRNSTRKSLKSPH-QNRLOEETNELDANVPPVSFLKVLKLNSTEWFFVVG	711	
QY	717	IFCAIINGLOPAFISIIERSIIGTFRDDEDPETKQNSMFSVFLVLGLIISFIFFLQ	776	

DB	712	TVCAIANGALQPAFSLIILSEMIAIFGPGDD--AVKQKCNMFLSLVLGLGLVLSFFTFQ	770	
QY	777	FTFGKAGEILTKRLRYMVFRLRQDVSWFDDPKNTTGTALTTRLANDAAQVKGALGSRIA	836	
DB	771	FTFGKAGEILTKRLRSMFAKMLRQDMSWFDHKNSTGALSTRLATDAAQVKGATGKLA	830	
QY	837	VITQNTANLGTGIIISLIYWGQITLIIILAIPIIAIAGVVMKMSGOALKDKKELEG	896	
DB	831	LIAQNTANLGTGIIISLIYWGQITLIIILAIPIIAIAGVVMKMSGOALKDKKELEG	890	
QY	897	KIATEAIENFRVWSLTREQEYMYAOSLOVPYRNSLRKAHIFGVSEITCAMMYFSYA	956	
DB	891	KIATEAIENIRVWSLTREQEYMYAOSLOVPYRNSLRKAHIFGVSEITCAMMYFSYA	950	
QY	957	GCFRGAYLVANEFMNFQDVLVFSALVFGAMAVQVSSFADYAKAVSAHVITWITEK	1016	
DB	951	GCFRGAYLVANEFMNFQDVLVFSALVFGAMAVQVSSFADYAKAVSAHVITWITEK	1010	
QY	1017	SLPDISYSPHGLKPNLTLEGNTFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGC	1076	
DB	1011	QPLDISYSGEGLWPKDFEGSVTFNEVFNYPTRANVPVLOGLSLEVKKGQTLALVSSGC	1070	
QY	1077	GKSTVVQLLERYDPLAGSVLDGKREIKHLNVQWLRHAGLIVSOBPIILFDCSIAENIAYG	1136	
DB	1071	GKSTVVQLLERYDPMAGSVLDGQBAKLNQWLRHAGLIVSOBPIILFDCSIAENIAYG	1130	
QY	1137	DNSRVVSHHEEIMQAAKEANIHFIETLPKYNTRYGDKGTQSLSGGKORAIARALVRQ	1196	
DB	1131	DNSRVVPHDEIVRAAKEANIHFIETLPKYNTRYGDKGTQSLSGGKORAIARALVRQ	1190	
QY	1197	HILLDEATSDALDTESEKVVQVALDKARGRTTIVIAHRLSTIQNALIVVFGKVKHEH	1256	
DB	1191	RVLLDEATSDALDTESEKVVQVALDKARGRTTIVIAHRLSTIQNALIVVFGKVKHEH	1250	
QY	1257	GTHQQLLAQKGYFSMVSVOAGAK 1280		
DB	1251	GTHQQLLAQKGYFSMVSVOAGAK 1274		
RESULT 9				
MDR3_CRIGR STANDARD; PRT; 1281 AA.				
AC	P23174;			
DT	01-NOV-1991 (Rel. 20, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Multidrug resistance protein 3 (P-glycoprotein 3).			
GN	PGY3 OR PGP3			
OS	Cricetulus griseus (Chinese hamster).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;			
OC	Cricetulus.			
OX	NCBI_TaxID=10029;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92135896; PubMed=1685679;			
RA	Endicott J.A., Sarangi F., Ling V.;			
RT	"Complete cDNA sequences encoding the Chinese hamster P-glycoprotein			
RT	gene family.";			
RL	DNA Seq. 2:85-101(1991).			
CC	-1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED			
CC	DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-1- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT			
CC	CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY			
CC	CANNOT.			
CC	-1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M60042; AAA68885.1; -
CC HSP; P13569; INBD.
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001140; ABC_transporter_tmem.
CC InterPro; IPR003439; ABC_transporter.
CC InterPro; IPR001687; ATP_GTP-A.
CC Pfam; PF00664; ABC_membrane; 2.
CC Pfam; PF00005; ABC_tran; 2.
CC SMART; SM00382; AAA; 2.
CC PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 299 319 POTENTIAL.
FT TRANSMEM 328 348 POTENTIAL.
FT DOMAIN 349 712 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 713 733 POTENTIAL.
FT TRANSMEM 758 778 POTENTIAL.
FT TRANSMEM 834 854 POTENTIAL.
FT TRANSMEM 855 875 POTENTIAL.
FT TRANSMEM 938 958 POTENTIAL.
FT TRANSMEM 975 995 POTENTIAL.
FT DOMAIN 996 1281 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 429 436 ATP (POTENTIAL).
FT NP_BIND 1071 1078 ATP (POTENTIAL).
FT SEQUENCE 1281 AA; 140866 MW; 2203EF61EBB29602 CRC64;

Query Match 74.5%; Score 4823.5; DB 1; Length 1281;
Best Local Similarity 73.7%; Pred. No. 7.5e-268;
Matches 949; Conservative 147; Mismatches 175; Indels 17; Gaps 8;

QY 1 MDPEGRKGSAREK-----NFKWKKKKKKKKKKPT-VSTFAFRYSNWLRLYLVLG 54
DB 1 MDLEAARNGTARRRGVTEGDFELSGISNQGRNKKKKKVLNGLPTLTFRYSDWQDDKLFLMG 60
QY 55 TMAAIIHGAALPLMLVFGNMTDSEA-NAGISRNKTFPVIINESTNTQTFINHLEEM 113
DB 61 TMAAIIHGAALPLMLVFGNMTDSEA-NAGISRNKTFPVIINESTNTQTFINHLEEM 113
QY 114 TTYAYTSGAGVLVAAYIQVSWFCLAGROILKIRKQFFHAIMRQELGFWHDVDEL 173
DB 114 TRYAYTSGAGVLVAAYIQVSWFCLAGROILKIRKQFFHAIMRQELGFWHDVDEL 173
QY 174 NTRLTDDVSKINEGIGDKYGMFFQSTATFTFTGTVGTRGWKLTLVLAISPVGLSAAI 233
DB 174 NTRLTDDVSKINEGIGDKYGMFFQSTATFTFTGTVGTRGWKLTLVLAISPVGLSAAI 233
QY 234 WAKILSFTDKELAYAKAGAAVEALAAIRTVIAFGQKKKELERYKNLEAKGIGIKK 293
DB 234 WAKILSFTDKELAYAKAGAAVEALAAIRTVIAFGQKKKELERYKNLEAKGIGIKK 293
QY 294 ATANISGAELLIIYASVALAFWGTSLVLSSEYSIGVLTVPFVSLIGAFSIGQASPS 353
DB 294 ATANISGAELLIIYASVALAFWGTSLVLSSEYSIGVLTVPFVSLIGAFSIGQASPS 353
QY 354 IEAFANARGAAVEIFKIDNKSIDSYSKSHKPNINLEFNKHFSPYSRKEVKILK 413
DB 354 IDAFANARGAAVEIFKIDNKSIDSYSKSHKPNINLEFNKHFSPYSRKEVKILK 413
QY 414 GLNLKVSQGTVALVNSGCGSTTTVQLMRLYDPTDGMVCIDGQDIRTNVRLHREITG 473
DB 414 GLNLKVSQGTVALVNSGCGSTTTVQLMRLYDPTDGMVCIDGQDIRTNVRLHREITG 473
QY 474 VVSQEPVLFAATTAEINIRYGRNVMTDETEKAVKANAYDFTMKLPNKFDTLVGERGQAL 533
DB 474 VVSQEPVLFAATTAEINIRYGRNVMTDETEKAVKANAYDFTMKLPNKFDTLVGERGQAL 533

DB 474 VVSQEPVLFAATTAEINIRYGRNVMTDETEKAVKANAYDFTMKLPNKFDTLVGERGQAL 533
QY 534 SGQKORIAIARALVNPKNLILDEATSDALTESEAVQVVALDKARKGRTTIVIAHRLST 593
DB 534 SGQKORIAIARALVNPKNLILDEATSDALTESEAVQVVALDKARKGRTTIVIAHRLST 593
QY 594 VRNADVIAGDDGVIIVEKGNHDELMEKGIYKFLVTMQTRNGNEIELEENATGSKSESAL 653
DB 594 VRNADVIAGDDGVIIVEKGNHDELMEKGIYKFLVTMQTRNGNEIELEENATGSKSESAL 653
QY 654 EMSPKDSGSLIKRRSTRSIIHAPQODKGLTKED-LNENPPVPSFWFRLKLNSTENWY 712
DB 654 GMTP-NGWKSHPFRNSTKSLKSLRAHHRDLVDDELDAINPPVPSFWFRLKLNSTENWY 712
QY 713 FVYGIFCAIINGLQPAFSIIFRIIGTIRDEDPETKQNSNMFSVLFLVIGISFIFF 772
DB 713 FVYGTVCIAVNGALQPAISILSEMIAIFGPGDD-AVKQKCNLFSLVFLGLGVLSFFTF 771
QY 773 FLOGFTFGKAGEILTRLRSMYPRSMRLQDVSWFDDPKNTGALTTRLANDAAQYKGAIG 832
DB 773 FLOGFTFGKAGEILTRLRSMYPRSMRLQDVSWFDDPKNTGALTTRLANDAAQYKGAIG 832
QY 833 SRLAVITONIANLGTGIISLIYGWLTLLLLAIIVPIIAIAGVEMKLSGQALKDKKEL 892
DB 833 SRLAVITONIANLGTGIISLIYGWLTLLLLAIIVPIIAIAGVEMKLSGQALKDKKEL 892
QY 893 EGAGKIATEAIENFRVSVLTREQKFEYMAQSLQVPYRNSLKAHIFGVSPSITQAMMY 952
DB 893 EGAGKIATEAIENFRVSVLTREQKFEYMAQSLQVPYRNSLKAHIFGVSPSITQAMMY 952
QY 953 FSYAGCFRFGAYLVANFNMFDVLLVFSIAIVGAMAVGVSSFPADYAKAKVSAAHVIM 1012
DB 953 FSYAGCFRFGAYLVANFNMFDVLLVFSIAIVGAMAVGVSSFPADYAKAKVSAAHVIM 1012
QY 952 FSYAGCFRFGAYLVANFNMFDVLLVFSIAIVGAMAVGVSSFPADYAKAKVSAAHVIM 1011
DB 952 FSYAGCFRFGAYLVANFNMFDVLLVFSIAIVGAMAVGVSSFPADYAKAKVSAAHVIM 1011
QY 1013 IIEKSPILDSYSPHGLKPNLTGNTFNEVFNYPDPIDVQLGLSLEVKKGOTLALVG 1072
DB 1013 IIEKSPILDSYSPHGLKPNLTGNTFNEVFNYPDPIDVQLGLSLEVKKGOTLALVG 1072
QY 1012 LFERQPLIDTSIEGLMPDKFEGSVTFNEVFNYPFRANMPVLOGLSLEVKKGOTLALVG 1071
DB 1012 LFERQPLIDTSIEGLMPDKFEGSVTFNEVFNYPFRANMPVLOGLSLEVKKGOTLALVG 1071
QY 1073 SSGCGKSTVVQLLRFYDPLAGSVLIDKEIKHLNVLWLAHILGIVSQEPILFDCSIAEN 1132
DB 1073 SSGCGKSTVVQLLRFYDPLAGSVLIDKEIKHLNVLWLAHILGIVSQEPILFDCSIAEN 1132
QY 1133 IAYGDNRRVVSHEIMQAAKEANHHFETLPEKYNTRVGVGDKGTQLSGGQKQRIARAL 1192
DB 1133 IAYGDNRRVVSHEIMQAAKEANHHFETLPEKYNTRVGVGDKGTQLSGGQKQRIARAL 1192
QY 1132 IAYGDNRRVVSHEIMQAAKEANHHFETLPEKYNTRVGVGDKGTQLSGGQKQRIARAL 1191
DB 1132 IAYGDNRRVVSHEIMQAAKEANHHFETLPEKYNTRVGVGDKGTQLSGGQKQRIARAL 1191
QY 1193 VROPHILLDEATSEKVVQEAALDKAREGTCIVIAHRLSTIQNADLIWTFQNGK 1252
DB 1193 VROPHILLDEATSEKVVQEAALDKAREGTCIVIAHRLSTIQNADLIWTFQNGK 1252
QY 1253 VKEHGTQQLLAQKGIYFSMVSVQAGAK 1280
DB 1253 VKEHGTQQLLAQKGIYFSMVSVQAGAK 1280
QY 1252 VKEHGTQQLLAQKGIYFSMVSVQAGAK 1279
DB 1252 VKEHGTQQLLAQKGIYFSMVSVQAGAK 1279

RESULT 10
MDR2_RAT STANDARD; PRT; 1278 AA.
AC Q08201;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 2 (p-glycoprotein 2).
GN PGY2 OR MDR2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBL_faxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISCHER;
RX MEDLINE=93376516; PubMed=8103593;
RA Brown P.C., Thorgeirsson S.S., Silverman J.A.;

DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11).
CN ABCB11 OR BSEP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT PFIC2 GLY-297.
RX MEDLINE=99021377; PubMed=9806540;
RA Strautnieks S.S., Bull L.N., Knisely A.S., Kocoshis S.A., Dahl N.,
RA Arnell H., Sokal E., Dahan K., Childs S., Ling V., Tanner M.S.,
RA Kagawalla A.F., Nemeth A., Pawlowska J., Baker A., Mieli-Vergani G.,
RA Freilmer N.B., Gardiner R.M., Thompson R.J.;
RT "A gene encoding a liver-specific ABC transporter is mutated in
RL progressive familial intrahepatic cholestasis.";
RL Nat. Genet. 20:233-238(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Mol O., Hooiveld G.J.E.J., Jansen P.L.M., Muller M.;
RT "Cellular localization and functional characterization of the human
RT bile salt export pump (BSEP)".
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
CC INTO THE CANALICULUS OF HEPATOCYTES.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
CC MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
CC SITU (BY SIMILARITY).
CC -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -!- DISEASE: DEFECTS IN ABCB11 ARE THE CAUSE OF PROGRESSIVE FAMILIAL
CC INTRAHEPATIC CHOLESTASIS 2 (PFIC2), AN INHERITED LIVER DISEASE OF
CC CHILDHOOD. PFIC2 IS CHARACTERIZED BY CHOLESTASIS AND NORMAL SERUM
CC GAMMA-GLOUTAMYLTRANSFERASE ACTIVITY.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF091582; AAC77455.1; -;
DR EMBL; AF136523; AAC28285.1; -;
DR MIM; 603201; -;
DR MIM; 601847; -;
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transmembrane; Transport; Disease mutation.
FT DOMAIN 1 62 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 63 83 POTENTIAL.
FT DOMAIN 84 147 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 148 168 POTENTIAL.
FT DOMAIN 169 215 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 216 236 POTENTIAL.
FT DOMAIN 237 240 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 241 261 POTENTIAL.
FT DOMAIN 262 319 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 320 340 POTENTIAL.
FT DOMAIN 341 353 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 354 374 POTENTIAL.
FT DOMAIN 375 755 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 756 776 POTENTIAL.

FT	DOMAIN	777	794	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	795	815	POTENTIAL.
FT	DOMAIN	816	869	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	870	890	POTENTIAL.
FT	TRANSMEM	891	911	POTENTIAL.
FT	DOMAIN	912	979	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	980	1000	POTENTIAL.
FT	DOMAIN	1001	1011	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1012	1032	POTENTIAL.
FT	DOMAIN	1033	1321	CYTOPLASMIC (POTENTIAL).
FT	NP_BIND	455	462	ATP (POTENTIAL).
FT	NP_BIND	1113	1120	ATP (POTENTIAL).
FT	CARBOHYD	109	109	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	122	122	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	297	297	E -> G (IN PFIC2).
FT				/FTid=VAR_010271.
FT	CONFLICT	339	339	V -> L (IN REF. 2).
FT	CONFLICT	444	444	V -> A (IN REF. 2).
FT	SEQUENCE	1321	AA; 146392	MW; D444CACC48DEA371 CRC64;
QY	Query Match	51.1%	Score 3305; DB 1; Length 1321;	
QY	Best Local Similarity	50.4%	Pred. No. 4.5e-181;	
QY	Matches	658; Conservative	242; Mismatches 366; Indels 40; Gaps 10;	
QY	8	KGSAEKFWKMGKSKKNEKK	-----EKK-----PTVSTFAMFRYSNMLDRILMLVGTMAAI	59
DB	11	KKFEENDGFESDKSYNNDKKRLQDEKGDGVRVGFQLFREFSSSTDILWLFVGSILCAF	70	
QY	60	IHGAAFLPMLVFGNMTDTSFANAGISR	-----NKT-----PVLINESITNNTOHF	105
DB	71	LHGIAQPGVLLIEFTMDYDFIDYDVELQEIQPKACVNNITVWNTSSLNQNTNGTRCG	130	
QY	106	INHLEEMTTVAYYSIGAGVVAAYIQVSFWCLAAGROILKIRKOFFHAIHQEIGWF	165	
DB	131	LLNTESEMIKFSYAGIYAVLITGYIQICFVIAAARQIKMRKFFPRIRMEIGWF	190	
QY	166	DVHDVDELNRLTDDVSKINEIGDKTGMFPQSITATFTTGFIVGTFRGWKLTLVLAISP	225	
DB	191	DCNSVGNLTFRSDIINKINDAIADQMAFLTQRTMTSTICGLGLGFFRGWKLTLVLSVSP	250	
QY	226	VLGLSAAIWKILSSFPDKELAYAKAGAAVEELAAIRVIAFGGOKKELERYKNLEE	285	
DB	251	LIGIGATIGLSVSKFYDYELKAYAKAGVVADEVISSRIVAAEFGGKREVEREKNLVF	310	
QY	286	AKGIGIKKAITANISIGAAFLLIYASYALAFWYGTSLVL-SSEYSIGQVLTFFSVLIGA	344	
DB	311	AQRWGIKRGIVMGFTGFVWCLIFLCYAVAFWYGSTLVLDGEYTPGTLVQIFLSVIGA	370	
QY	345	FSIGQASPSIEAFANAGAAYEIFKIDNKPISDYSKSGHKPDNIKGNLEFKNVHSYP	404	
DB	371	LNLGNASPCLEAFATGRAATSIFETIDRKPIDDCMEDSGYKLDRIKGEIEFHNVTTHYP	430	
QY	405	SRKEVKILKGLNKLKVSQGTVALVNSGCGKSTVQLMORLYDPTDGMVCIDGQDITIN	464	
DB	431	SRPEVKILNLMVMKPGEMTALVPGSGAKSTALQIQFYDPCGMVVDGHDHISLN	490	
QY	465	VRLHREITGVYSQBPVLFATTIAENIRYGRNVTMDIEKAVKEANAYDFITMKLPNKFT	524	
DB	491	IQLWRDQIGIVEQEPVLFSTTIAENIRYGRNVTMDIEKAVKEANAYDFITMKLPNKFT	550	
QY	525	LVGERGALSGGQKQRTAIARALVRNPKILLIDATSEALDSEAVVQVALDKARKGRTT	584	
DB	551	LVGEGGQMGSGGQKQRTAIARALVRNPKILLIDATSEALDSEAVVQVALDKARKGRTT	610	
QY	585	IVIAHRLSTVRNADYVAGFDGVIKGNHDELAKRGYFKLVMTQTRG-----NEIELE	640	
DB	611	ISVAHRLSTVRNADYVAGFDGVIKGNHDELAKRGYFKLVMTQTRG-----NEIELE	670	
QY	641	NATGESKSEDALEMPKDSGSSIIKRRSTR-----SIHAP--QGDRKLGTGED-----	689	
DB	671	DATEDDDMLARTFSRGSYQDSLRASRQRKSQLSVLVHPPPLVAVVDHKSVEEDRKDKDI	730	


```
Db 270 ELKAVAGSVADEISSMRTVAAGGKKEVEREKNLVFAQRWGRKGVMGFTGYM 329
Qy 305 FLIIYASYALATWYGTSLVL-SSEYSIGOVLTVPFVSVLIGAFSICQASPSIEAFANARGA 363
Db 330 WCLIFCYALAFWYGSKLIVGEYSFGALVQIFLSVIGALNLGNASPCLEAFAGARGA 389
Qy 364 AYEIEFKIDNKPSIDSYSKSGHKPNKGNLFKKNVHFSYPSRKEVKTLKGLNLKLVQSGQ 423
Db 390 ASSIFETIDRKPIIDCMSEGDGKLEKRIEIEFHNVTHTYPSRPEVKLNLSWIKPGE 449
Qy 424 TVALVNSCGKSTTVQLMQRLYDTDGMVCIDGDGDIRTINVRHLREITGVVYSQBPVLFA 483
Db 450 MVALVGPAGKSTALQLLHRYFGTEGMVTVESHDIRSSHQIQLRNLQIGIVEQEPVLFF 509
Qy 484 TTAENIRYGRNVMTDETEKAVKANAYDFIMKLPNKFDILVGERGAQLSGGQKQRTAI 543
Db 510 HTAERIRYGRDATMEDLIQAAKEANAYNFTMDLPQOQFDTLVGGGQMSGGQKQRTAI 569
Qy 544 ARALVRNPKILLDDATSAIDTSEAVVQVADLKARKGRTTIVIAHRLSTVRNADVIAGF 603
Db 570 ARALVRNPKILLDDATSAIDTSEAVVQVADLKARKGRTTIVIAHRLSTVRNADVIAGF 629
Qy 604 DGVYIVKEGNHDELMEKGIYFKLVMTQTRGNEIELE-----NATGESKESDALEMSPKD 659
Db 630 EHGAVERGTEBELERKGVYALVTLQSORNQDQEEENEKDATEDDIPKFTSRGNQOD 689
Qy 660 SGSSLLKRS-----TRSHIAPQODQRKLTGKEDLNENVPVPSFWRL 703
Db 690 SLRASLRQKRSQLSYLAHEPPMAVEDHKSTHEDRKQDLPAQED----IEPASVRRIM 745
Qy 704 KLNSTWMPYVVGIFCAIINGLQPAFSIIFSRIGIFTRDDEPTKRONSMFVFLV 763
Db 746 KLNAPWYLLGSMCAVNGAVTPYAFLSQILGTSL-PDKEQRQINGICLLFVT 804
Qy 764 LGHISFTFFLOGFTGKAGEILTKRLRYMVRMSLRQDVSWFDDPKNTTGALTTIRLAND 823
Db 805 LGCVSFTQFLOGYTFKAGSGLLTKRLKRGFRAMLQDQIGWFDLRLNSPGALTTIRLATD 864
Qy 824 AAQVKAIGSRILAVITQNTANLGTGIIISLIYQWLTLLLAIVPIATAGVEMKMLSG 883
Db 865 ASVQVATGSGQGMVNSTVNTVAMIAPLFSWKLITGIVCFPLFALSGALQTKMLTG 924
Qy 884 QALKDKKELEGAGKATEAIEENFRVTVSLTRQKFEYMYAQSLQVPYRSLRKAHIFGVS 943
Db 925 FASRDKALEKAGQITSEALSNIITVAGIGKERKFTETFEAELEKPYKMAIKANVYGLC 984
Qy 944 FSITQAMVFSYAGCPRFGAYLVANEFMFQDVLVFSIAIVFGAMAVGVQVSSFADPYAKA 1003
Db 985 FGFSGCITETANSASYRGYGLISNGLHFSYVFRVISAVILSATALGRASSYTPSYAKA 1044
Qy 1004 KVSAAHVIMIEKSLPIDSYSPHGLKPNLTLEGNVTNEVFNVPTRPDIPVLOGLSLEVK 1063
Db 1045 KISAAFFQLDQPPINIVYSSAGEKWDNFQKIDFVDCKFYTPSRPDQVNLGLSVMS 1104
Qy 1064 KGOTLALVSGSGCKSTVQVLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHGIYSQEP 1123
Db 1105 PRQTLAFVSGSGCKSTSIOLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHGIYSQEP 1164
Qy 1124 LFCSTAEATAYADNSRVSVSHIEIMQAAKEANTHHFETLPEKYNTVRGDKGTQLSGGOK 1183
Db 1165 LFACSKDKNIKYGDNTQETIPMERITAAAKQAQVHDVFMVSLPEKYNTVRGDKGTQLSGGOK 1224
Qy 1184 ORTAIRALVRQPHILLDEATSAIDTSEAVVQVADLKARKGRTTIVIAHRLSTVRNADVIAGF 1243
Db 1225 ORTAIRALVRQPHILLDEATSAIDTSEAVVQVADLKARKGRTTIVIAHRLSTVRNADVIAGF 1284
Qy 1244 LIYVFGNGKVEHGTHTQQLAQAQGIYFSMVS 1274
Db 1285 IIAVMSQGVIEKGTHTHEELMVQKGYKLV 1315
```

RESULT 13

```
AB11_MOUSE
ID AB11_MOUSE STANDARD; PRT: 1321 AA.
AC Q90Y30; Q90Z8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11)
DE (Sister of P-glycoprotein).
GN ABCB11 OR BSEP OR SPGP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=20076398; PubMed=10607905;
RA Green R.M., Hoda F., Ward K.B.;
RT "Molecular cloning and characterization of the murine bile salt export
pump.";
RL Gene 241:117-123(2000).
RN [2]
RP SEQUENCE OF 463-635 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RA Salkar R., Suchy F.J., Ananthanarayanan M.;
RT "Molecular cloning of mouse liver bile salt export pump (bsep).";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
CC INTO THE CANALICULUS OF HEPATOCYTES.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
CC MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
CC SITU.
CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: AF133903; AAF14372.1; -.
DR EMBL; AF186585; AAD56419.1; -.
DR MGD; MGI:1351619; Abcb11.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmnm.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_cran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR ATP-binding; Transmembrane; Transport.
KW DOMAIN 1 62 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 63 83 POTENTIAL.
FT DOMAIN 84 147 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 148 168 POTENTIAL.
FT DOMAIN 169 215 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 216 236 POTENTIAL.
FT DOMAIN 237 240 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 241 261 POTENTIAL.
FT DOMAIN 262 319 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 320 340 POTENTIAL.
FT DOMAIN 341 353 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 354 374 POTENTIAL.
FT DOMAIN 375 755 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 756 776 POTENTIAL.
```


FT DOMAIN 777 794 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 795 815 POTENTIAL.
 FT DOMAIN 816 869 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 870 890 POTENTIAL.
 FT TRANSMEM 891 911 POTENTIAL.
 FT TRANSMEM 912 979 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 980 1000 POTENTIAL.
 FT TRANSMEM 1001 1011 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1012 1032 POTENTIAL.
 FT TRANSMEM 1033 1321 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 455 462 ATP (POTENTIAL).
 FT NP_BIND 1113 1120 ATP (POTENTIAL).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 481 481 L -> P (IN REF. 2).
 FT CONFLICT 633 633 T -> V (IN REF. 2).
 SQ SEQUENCE 1321 AA; 146675 MW; 15B5EBF175D32967 CRC64;

 Query Match 49.6%; Score 3212.5; DB 1; Length 1321;
 Best Local Similarity 49.7%; Pred. No. 8.7e-176;
 Matches 643; Conservative 237; Mismatches 376; Indels 37; Gaps 9;

 QY 15 FWKMGKSK-KNEKKEKPTVSTFAMFRYNSWLDRLVLTMAAIIHGAALPLMLVFG 73
 DB 25 FHNDDKSKRLQDKKKGEGARVPPELFRSSKDNWLMFNGSVCALLHGAQPMIIVFG 84

 QY 74 NMTDSPANAGISR-----NKTFFPVI---INSITNNQHFNIHLEEMTYAYY 119
 DB 85 ILTDIFVEYDIEROELSIPGKVCNMNTIVINSFNOMTNGTSCGLVDINSEVIFKPSGI 144

 QY 120 YSGIGAGVLAAYIQVSWFCLAGROILKTKOFFHAIMROETGWDVHDVGLNRLTD 179
 DB 145 YAGVGAVILGVOIRLWITGARQIRKMKFFIRIMREIGWEDCTVSGLNSRFSF 204

 QY 180 DVSKINGIGDKIMGFQSIATFTFTGIVGTRGKWLTVILAIISPVILGSAIAWAKILS 239
 DB 205 DINKIDEAIDQMALFLQRLSTALSGLLGYRGWKLTVILAVSPILIGIAVIGLSVA 264

 QY 240 SFTDKELLAYAKAGAAVEVLAIRTVIAFGQKKELERYKNKLEAKGIGIKKATANI 299
 DB 265 KFTLEUKAKAGSIADEVLSSIRTVAAFGGENKEYERKMLFAQRWIKMGVMGF 324

 QY 300 SIGAAFLLIYASALAPWYGTSLVL-SSEYSIGOVTVFVSLVIGAFSIGOASPSIEAFA 358
 DB 325 FTGVMMCLIFCYALAFWYGRSLVLDGEYPTGLIQIFLCVIAAMNIGNASSCLEIFS 384

 QY 359 NARGAAEIPKIIDNKPISIDYSKSGHKPDNIKGNLEFKNVHFSYPSRKEVKILGNLK 418
 DB 385 TGCSSAASSIFQTIIDROPVMDCMGDKYKLDRIKGEIEFHNVTTHYPSRPEVKILNLSV 444

 QY 419 VQSGOTVALVNSGCKSTTVOLMQRLYDPTDGMVCDIGODIRTNVRLHREITGVVSOE 478
 DB 445 IKPGETAFVSSGAGKSTALQIIRFYDPCGEMVTLGDHDIRSLNRLDQIGIVEOE 504

 QY 479 PVLFAITIAENIRYGRNVTMDIEKAVKANAYDFIMKLPNKFDTLVGERAQLSGGOK 538
 DB 505 PVLFTTIAENIRLGREATMEDIVQAKADANAYNFIMALPOQDFLVGEGGQMSGGOK 564

 QY 539 ORTAIRALVNPKILLDDATSDALDTESEAVVQVQALDKARKGRTTIVIAHRLSTVRNAD 598
 DB 565 QRYAIAIRALIRKPKILLDDMATSDALDNESEAKVOGALNKIQHCHTIIISVAHRLSTVRSAD 624

 QY 599 VIAGFDGVIYVEKNDELKMKGIYFKLVTWOTRNELENA-TGESKSSSDALEM-- 655
 DB 625 VIIGFEGHTAVERGTHEELERKGVYFVLTLQSQEDNTHKETGKDKTGTGDPERTF 684

 QY 656 ---SPKDSGSSLLKRRSTR-----SIHAPQODRKLGFKEDLNENVPVPSFW 700
 DB 685 SRGSYQDSLRAIRQRKSKQLSHLSHEPPLATGDHKSSEDRK--DNDVLVEVEPAPVR 742
 QY 701 RILKLNSTWPFVVGIFCAIINGGLQPAFSIFSRILGIIFTRDEDPETKQNSNMFSVL 760

DB 743 RILKYNISEPPIYILVGAALCAANGAVTPIYSLLFSQILKTFSL-VDKEQORSEIYSMCLF 801
 QY 761 FLVLGIISFTFFLQGTFTFGKAGEILTCLRMYMVFMSMLRQDVSFDDPKNTTGALTTRL 820
 DB 802 FVILGCVSLFTQFLOGYNFAKSGELLTKRLRKFGFKAMLRODQIGWEDDLKNNPVLITRL 861
 QY 821 ANDAAQVKGAGSLRLAVITONIANLGTGIIISLIYQWQTLTLLAIVPIAIAGVVEMK 880
 DB 862 ATDASQVQATGSOVGMVMNSFTNIEFVAVLIAFLFNKLSLVISFPFELALSGAVQTKM 921
 QY 881 LSGOALKDKLEKAGKIAETENFRTVTVSLTREKFEYMAQSLQVYRNSLRKAHIF 940
 DB 922 LTGFSQDKLEIKAKQIYNEALNSRTVAGICVEGRFKAFEVELEKSYKTAIRKANVY 981
 QY 941 GVSFSTQAMMYSYACGFRFGAYLVANBFMFQDVLVLSAIVFGAMAVGVSSFADPY 1000
 DB 982 GLCYAFSGISFLANSAAYRYGGYLIVYEDLNFYSYFVRVSSIAMSATAVGRTFSYTPSY 1041
 QY 1001 AKAKVSAAHVMIIEKSPILDSYSPHGLKPNLTLEGNTVNEVFNPTPRDIPVQLGSL 1060
 DB 1042 AKAKISAAARFFQLDRKPPIDVYSGAGEKWNDFQKIDFIDCKFTYPSRPDIQVNLGLSV 1101
 QY 1061 EVKKGOTLALVGSSGCGKSTVVQLLRFERYDPLAGSVLIDGKEIKHLNWOHLRAHLGIYSQ 1120
 DB 1102 SVDPGOTLAFVSSGCGCKSTISQLLERFYDPOQGVIMIDGHSKKNVVOFLKSNIGIYSQ 1161
 QY 1121 EPILFPCSAENIATYGDNSRVVSHSEEMOAKAEANIHFETLPKYNTRVGDGKTQSLG 1180
 DB 1162 EPVLFPCSIMDKYDNTKEISVERAIAAQAQLHDFVMSLPKEYETNVGIQGSQSLR 1221
 QY 1181 GQKQRTAIRALVRQPHIILLDEATSDALDTESEKVVQEAALDKAREGRTICIVIAHRLSTIQ 1240
 DB 1222 GEKQRTAIRAIRVROPKILLDEATSDALDTESEKTVQALDKAREGRTICIVIAHRLSTIQ 1281
 QY 1241 NADLIWVFGNGKHEGTHQOOLLAOKGIYFSMW 1273
 DB 1282 NSDIATVMSQGVVIEKGTGKHLMDQKGYKLV 1314

 RESULT 14
 AB11_RAT
 ID AB11_RAT STANDARD; PRT; 1321 AA.
 AC 070127;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11)
 DE (Sister of P-glycoprotein).
 GN ABCB11 OR BSEP OR SPGP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
 RX MEDLINE=98212048; PubMed=9545351;
 RA Gerloff T., Stieger B., Hagenbuch B., Madon J., Landmann L., Roth J.,
 RA Hofmann A.F., Meier P.J.;
 RT "The sister of p-glycoprotein represents the canalicular bile salt
 RT export pump of mammalian liver.";
 RL J. Biol. Chem. 273:10046-10050(1998).
 CC -|- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
 CC INTO THE CANALICULUS OF HEPATOCYTES.
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -|- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
 CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
 CC MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
 CC SITU.
 CC -|- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
 CC ATP BINDING CASSETTE (ABC) DOMAIN.

[illegible]

Search completed: November 6, 2002, 18:40:34
Job time : 17.4981 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:31:39 ; Search time 29.3946 Seconds
(without alignments)
7539.009 Million cell updates/sec

Title: US-09-672-725c-23
Perfect score: 6474
Sequence: 1 MDPGGRGSAEKNFWMGK.....LLAQGIYFSVSVQACAKR 1281

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL19.*
1: sp.archaea.*
2: sp.bacteria.*
3: sp.fungi.*
4: sp.human.*
5: sp.invertebrate.*
6: sp.mammal.*
7: sp.mhc.*
8: sp.organelle.*
9: sp.phage.*
10: sp.plant.*
11: sp.rodent.*
12: sp.virus.*
13: sp.vertbrate.*
14: sp.unclassified.*
15: sp.virus.*
16: sp.bacteriap.*
17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6425.5	99.3	1280	6 046605	046605 canis famil
2	5687.5	87.9	1285	6 002793	002793 ovis aries
3	5634	87.0	1272	11 09JK64	09JK64 rattus norv
4	5612	86.7	1272	11 0924L6	0924L6 rattus norv
5	5521.5	85.3	1163	6 09TSU2	Q9tsu2 felis silve
6	5338	82.5	1169	11 060502	060502 cricetus
7	4665.5	72.1	1288	13 093437	093437 gallus gall
8	4435	68.5	1287	13 091586	091586 xenopus lae
9	3336	51.5	1348	13 090235	090235 raja erinac
10	3214.5	49.7	1321	11 088331	088331 rattus norv
11	2862	44.2	1294	5 045721	045721 caenorhabdi
12	2832	43.7	1265	5 001495	001495 caenorhabdi
13	2825	43.6	851	13 09W693	09W693 fundulus he
14	2823.5	43.6	1275	5 061301	061301 haemochus
15	2746	42.4	1321	5 021349	021349 caenorhabdi
16	2693	41.6	1313	5 09V626	09V626 drosophila

17	2644	40.8	1283	5 024393	024393 drosophila
18	2612.5	40.4	817	13 090WV6	090WV6 pseudopleur
19	2608.5	40.3	1285	10 09LX1	09LX1 oryza sativ
20	2607.5	40.3	1286	10 080725	080725 arabidopsis
21	2601	40.2	1292	10 09ML09	09ML09 arabidopsis
22	2594.5	40.1	1278	10 09FWX7	09FWX7 arabidopsis
23	2555	39.5	1279	5 09V616	09V616 drosophila
24	2533	39.1	1348	3 09Y8G1	09Y8G1 emeritella
25	2526	39.0	1229	10 09SYI2	09SYI2 arabidopsis
26	2523	39.0	1229	10 049749	049749 arabidopsis
27	2523	39.0	1230	10 09SYI3	09SYI3 arabidopsis
28	2522	39.0	1289	10 094IH6	094IH6 coptis japo
29	2512	38.8	1349	3 043121	043121 aspergillus
30	2478	38.3	1229	10 09FWX8	09FWX8 arabidopsis
31	2467	38.1	1292	3 09C163	09C163 rhizomucor
32	2465.5	38.1	1313	10 09ZRG2	09ZRG2 solanum tub
33	2446	37.8	1408	3 043140	043140 cryptococcu
34	2442.5	37.7	1323	10 09W0M2	09W0M2 arabidopsis
35	2428	37.5	1310	5 024852	024852 entamoeba h
36	2424.5	37.4	1248	10 09FHF1	09FHF1 arabidopsis
37	2423	37.4	1286	10 09ZRW2	09ZRW2 trichophyto
38	2419.5	37.4	1331	3 09HGT5	09HGT5 caenorhabdi
39	2407	37.2	1266	5 020331	020331 caenorhabdi
40	2402	37.1	1252	10 09LJX0	09LJX0 arabidopsis
41	2398	37.0	1268	5 020335	020335 caenorhabdi
42	2385	36.8	1302	5 024851	024851 entamoeba h
43	2383	36.8	1320	5 09VRW2	09VRW2 drosophila
44	2352	36.3	538	11 003982	003982 cricetus
45	2335.5	36.1	1222	10 09SVW6	09SVW6 arabidopsis

ALIGNMENTS

RESULT 1
046605 PRELIMINARY; PRT; 1280 AA.
ID AC 046605;
AT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DI 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE P-GLYCOPROTEIN.
GN MDR1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Fuel O., Lepage J.F., Alvinerie M., Galtier P., Pineau T.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF045016; AAC02113.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR Pfam; PF00664; ABC_membrane; 2.
DR SMART; SM00382; ABC_tran; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
SQ ATP-binding; Transport.
KW SEQUENCE 1280 AA; 141524 MW; 762DD5AFF4C73306 CRC64;

Query Match 99.3%; Score 6425.5; DB 6; Length 1280;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1273; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 1 MDPGGRGSAEKNFWMGKSKKKEKKKKTSTVFAMFRYSNWLDRMLVGVMAAII 60
DB 1 MDPGGRGSAEKNFWMGKSKKKEKKKKTSTVFAMFRYSNWLDRMLVGVMAAII 59

```
QY 61 HGAALPLMLLVFGNMTDFANAGISRKTFPVIINESITNNTOHFINHLEEMTYAYYY 120
|||||
Db 60 HGAALPLMLLVFGNMTDFANAGISRKTFPVIINESITNNTOHFINHLEEMTYAYYY 119
|||||
QY 121 SGIGAGVLAAYIOVSFWCLAAAGROILKIRKOFFHAIHQGFHIMRQIGFVDHVDGELNRLTDD 180
|||||
Db 120 SGIGAGVLAAYIOVSFWCLAAAGROILKIRKOFFHAIHQGFHIMRQIGFVDHVDGELNRLTDD 179
|||||
QY 181 VSKINEGIDKIGMFQSIATFTFGIVFTRGWKLTLVILAIISPVLGLSAAIWAAILSS 240
|||||
Db 180 VSKINEGIDKIGMFQSIATFTFGIVFTRGWKLTLVILAIISPVLGLSAAIWAAILSS 239
|||||
QY 241 FTDKELLAYAKAGAAVEVLAAIRTVIAPFGQKKELEKRYNKNLEAKGIGIKKAITANIS 300
|||||
Db 240 FTDKELLAYAKAGAAVEVLAAIRTVIAPFGQKKELEKRYNKNLEAKGIGIKKAITANIS 299
|||||
QY 301 IGAAPLLIYASALAFWYGTSLVLSSEYSIGQVLTVPFVSLVIGAFSTGQASPIEAFANA 360
|||||
Db 300 IGAAPLLIYASALAFWYGTSLVLSSEYSIGQVLTVPFVSLVIGAFSTGQASPIEAFANA 359
|||||
QY 361 RGAAYEIKIIDNKPSIDSYKSGHKPDNIKGNLEFNHPSPSRKEVKILKGLNKVQ 420
|||||
Db 360 RGAAYEIKIIDNKPSIDSYKSGHKPDNIKGNLEFNHPSPSRKEVKILKGLNKVQ 419
|||||
QY 421 SGTVALVNSGCGKSTTVQLMORLYDPTDGMVCIIDGQDRTINVRHLREITGVVSOBVP 480
|||||
Db 420 SGTVALVNSGCGKSTTVQLMORLYDPTDGMVCIIDGQDRTINVRHLREITGVVSOBVP 479
|||||
QY 481 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGALSGGQKOR 540
|||||
Db 480 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGALSGGQKOR 539
|||||
QY 541 IAIARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
|||||
Db 540 IAIARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 599
|||||
QY 601 AGFDGCVIVEGNHDELMKEGIFYKLYVTMQTRGNEIENENATGESKSDALENSPKDS 660
|||||
Db 600 AGFDGCVIVEGNHDELMKEGIFYKLYVTMQTRGNEIENENATGESKSDALENSPKDS 659
|||||
QY 661 GSSLKRRSTRSHPAQGDRLKGTEDLNENPVVPSFWRILKLNSTEMPVYVVGIFCA 720
|||||
Db 660 GSSLKRRSTRSHPAQGDRLKGTEDLNENPVVPSFWRILKLNSTEMPVYVVGIFCA 719
|||||
QY 721 IINGLOPAFIIISRIIGITFTRDEDPETKQNSNMFSLFLVLGIISITTFLOGFTFG 780
|||||
Db 720 IINGLOPAFIIISRIIGITFTRDEDPETKQNSNMFSLFLVLGIISITTFLOGFTFG 779
|||||
QY 781 KAGEILTKRLRYMFRSMLRQDVSFDDPKNTTGALTTRLANDAAOVKGAIGSLRAVITQ 840
|||||
Db 780 KAGEILTKRLRYMFRSMLRQDVSFDDPKNTTGALTTRLANDAAOVKGAIGSLRAVITQ 839
|||||
QY 841 NTANLGTGIIISLYGHOLTLILLIATVPIIATAGVEMKMLSGQALKDKKELEGAGKIAT 900
|||||
Db 840 NTANLGTGIIISLYGHOLTLILLIATVPIIATAGVEMKMLSGQALKDKKELEGAGKIAT 899
|||||
QY 901 EAIENFRVSLTRQKFEYMAQSLQVPYRNSLRKAHIFCVSFSITOAMMYFSYAGCFR 960
|||||
Db 900 EAIENFRVSLTRQKFEYMAQSLQVPYRNSLRKAHIFCVSFSITOAMMYFSYAGCFR 959
|||||
QY 961 FCAYLVANEFNFDVLLVFSATVFCGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPLI 1020
|||||
Db 960 FCAYLVANEFNFDVLLVFSATVFCGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPLI 1019
|||||
QY 1021 DSYSPHGLKPTNLEGNTVNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVGSSCGKST 1080
|||||
Db 1020 DSYSPHGLKPTNLEGNTVNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVGSSCGKST 1079
|||||
QY 1081 VVOLLERYDPLAGSLVLDGKEIKHLNVQWIRAHLGIVSOEPIFPCSTAEINAYGDNSR 1140
|||||
Db 1080 VVOLLERYDPLAGSLVLDGKEIKHLNVQWIRAHLGIVSOEPIFPCSTAEINAYGDNSR 1139
|||||
QY 1141 VVSHEETMQAAKEANIHHFIETLPEKYNTRYGDKGTQLSGGQKORIAIARALVRQPHILL 1200
|||||
```

```
Db 1140 VVSHEETMQAAKEANIHHFIETLPEKYNTRYGDKGTQLSGGQKORIAIARALVRQPHILL 1199
|||||
QY 1201 LDATSALETSEKVVQBALOKAREGRTCVIAHRLSTONADLIIVFQNGKVKKEGHTHQ 1260
|||||
Db 1200 LDATSALETSEKVVQBALOKAREGRTCVIAHRLSTONADLIIVFQNGKVKKEGHTHQ 1259
|||||
QY 1261 QLLAOKGIIYFSMVSYOAGAKR 1281
|||||
Db 1260 QLLAOKGIIYFSMVSYOAGAKR 1280
|||||
RESULT 2
002793 PRELIMINARY; PRT; 1285 AA.
ID 002793;
AC 002793;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN-1.
DI MORI.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Longley M., Crawford A.M.;
RT "Ovine mdr1 gene.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC EMBL: U786609; AAB58489.1; -.
CC InterPro: IPR003593; AAA_transporter_tmern.
CC InterPro: IPR001140; ABC_transporter.
CC InterPro: IPR003439; ABC_transportfr.
CC Pfam: PF00664; ABC_membrane; 2.
CC Pfam: PF00005; ABC_tran; 2.
CC SMART: SM00382; AAA; 2.
CC PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ
SEQUENCE 1285 AA; 142020 MW; 90153EG6I7C44856F CRC64;

Query Match 87.9%; Score 5687.5; DB 6; Length 1285;
Best Local Similarity 87.4%; Pred. No. 0;
Matches 1123; Conservative 80; Mismatches 77; Indels 5; Gaps 3;

QY 1 MDPEGGKRG-SAEKNFWKMKK---SKNKKKKKPTVSTFAMPRYSNWLDRYMLVGTM 56
|||||
Db 1 MDLEGDRNRAAGGNFLKDKKRFSSKKDEKKKRTVSTFMFRYSNWLDRCLWLGTL 60
|||||
QY 57 AAIHGAALPLMLLVFGNMTDSPANAGISRKTFPVIINESITNNTOHFINHLEEMTY 116
|||||
Db 61 AAIHGAALPLMLLVFGNMTDSPANAGISRKTFPVIINESITNNTOHFINHLEEMTY 119
|||||
QY 117 AYTYSIGAGVLAAYIOVSFWCLAAAGROILKIRKOFFHAIHQGFHIMRQIGFVDHVDGELNTR 176
|||||
Db 120 AYTYSIGAGVLAAYIOVSFWCLAAAGROVHRIRKOFFHAIHQGFHIMRQIGFVDHVDGELNTR 179
|||||
QY 177 LTDVSKINEGIDKIGMFQSIATFTFGIVFTRGWKLTLVILAIISPVLGLSAAIWA 236
|||||
Db 180 LTDVSKINEGIDKIGMFQSIATFTFGIVFTRGWKLTLVILAIISPVLGLSAAIWA 239
|||||
QY 237 ILSSTFDKELLAYAKAGAAVEVLAAIRTVIAPFGQKKELEKRYNKNLEAKGIGIKKAIT 296
|||||
Db 240 ILSSTFDKELLAYAKAGAAVEVLAAIRTVIAPFGQKKELEKRYNKNLEAKGIGIKKAIT 299
|||||
QY 297 ANTISGAAPLLIYASALAFWYGTSLVLSSEYSIGQVLTVPFVSLVIGAFSTGQASPIEA 356
|||||
Db 300 ANTISGAAPLLIYASALAFWYGTSLVLSSEYSIGQVLTVPFVSLVIGAFSTGQASPIEA 359
|||||
QY 357 FANARGAAEYFKIIDNKPSIDSYKSGHKPDNIKGNLEFNHPSPSRKEVKILKGLN 416
|||||
```

OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Liver;
 RA Hooiveld G.J.E.J., Wilms J.W.J., Hagenbuch B., Jansen P.L.M.,
 RA Meijer D.K.F., Muller M.;
 RA "Cloning and functional characterization of the rat multidrug
 RT resistance protein Mr1a."; to EMBL/GenBank/DBJ databases.
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL: AF257746; AAF69007.1; -;
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001140; ABC transporter_tmern.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00664; ABC_membrane; 2.
 DR Pfam: PF00005; ABC_tran; 2.
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
 DR ATP-binding; Transposon.
 KW ATP-binding; Transposon.
 SQ SEQUENCE 1272 AA; 140326 MW; B5C1DF89DFC986C5 CRC64;
 Query Match 87.0%; Score 5634; DB 11; Length 1272;
 Best Local Similarity 86.5%; Pred. No. 1.4e-316;
 Matches 1108; Conservative 80; Mismatches 83; Indels 10; Gaps 2;
 QY 1 MDPSGRKGSAAKFNKMGKSKKKEKPTVSTFAMFYSNWLDRYMLVGTMAAII 60
 DB 1 MELEDLNGRADKNFSKMGKSKK-EKKEKPAVSVLTMPFYAGWLDRFYMLGLTAAII 59
 QY 61 HGAALPLMLVFGNMTDSFANAGISRNKTFPVVINESITNTQHFVNHLEPEETVYVY 120
 DB 60 HGIALPLMLVFGDMTDSFANVGNRSMSE-----YNTDIIYAKLEDEMTYVYV 110
 QY 121 SGIGAGVLVAAYIOVSFWCLAAAGQILKIRKQFFHAIMRQIEGDFVHDVGEINLRLTDD 180
 DB 111 TGIGAGVLIVAYIOVSLWCLAAAGQIHKIRKQFFHAIMRQIEGDFVHDVGEINLRLTDD 170
 QY 181 VSKINEGDKIGMFFQSIATFFGFIIVGTFGRGKLTIVILASPVGLSAAIWAKILSS 240
 DB 171 VSKINEGDKIGMFFQAMATFFGGFIIGFTGRGKLTIVILASPVGLSAAIWAKILSS 230
 QY 241 FTDKELLAYAKAGAAEVLAAIRTVIAFGGKKELEKRYNNLEKAEKIGIKKAITANIS 300
 DB 231 FTDKELQAYAKAGAAEVLAAIRTVIAFGGKKELEKRYNNLEKAEKIGIKKAITANIS 290
 QY 301 IGAALFLLIYASVALAFWGTSLVLSSEYSIGOVLTVPFSVLIGAFSGIQAASPIEAFANA 360
 DB 291 MGAALFLLIYASVALAFWGTSLVLSKEYTIGOVLTVPFSVLIGAFSGIQAASPIEAFANA 350
 QY 361 RGAAYEIFKIIDNKPISIDSYSGHKPDKNIQNLKFNHFSYPSRKEVKILGLNLKVQ 420
 DB 351 RGAAYEVFSIIDNKPISIDSYSGHKPDKNIQNLKFNHFSYPSRKEVKILGLNLKVQ 410
 QY 421 SGOTVALVNSGCGKTTVQLMQRLYDPTDGMVCIQDQIRTNVYHRLREITGVVSOBPV 480
 DB 411 SGOTVALVNSGCGKTTVQLMQRLYDPTDGEVSIQDQIRTNVYHRLREITGVVSOBPV 470
 QY 481 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGRGAQLSGGQKOR 540
 DB 471 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGRGAQLSGGQKOR 530
 QY 541 IATARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGTTIVIAHRLSTVRNADVI 600
 DB 531 IATARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGTTIVIAHRLSTVRNADVI 590
 QY 601 AGPDGIVVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKSSDALEWSPKDS 660
 DB 591 AGPDGIVVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKDGIDVNDWSSKDS 650
 GN PGY1.


```

OX NCBI_TaxID=10031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=91154265; PubMed=1671863;
RA Devine S.E., Hussain A., Davide J.P., Melera P.W.;
RT "Full length and alternatively spliced pgp-1 transcripts in multidrug
RT resistant Chinese hamster lung cells.";
RL J. Biol. Chem. 266:4545-4555(1991).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; M59254; AAA37005.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmam.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1169 AA; 128939 MW; 72E25B7CE29DC185 CRC64;

Query Match 82.5%; Score 5338; DB 11; Length 1169;
Best Local Similarity 89.3%; Pred. No. 1.4e-299;
Matches 1044; Conservative 69; Mismatches 56; Indels 0; Gaps 0;

QY 113 MTTAYYYSGIGAGLVAAIYQVSWCLAAAGROILKIRKOFFHAFHAIIMROEIGWFDVHDVGE 172
DB 1 MTTAYYYTGIGAGLVIAIYQVSWCLAAAGROIHKIRKOFFHAFHAIIMROEIGWFDVHDVGE 60

QY 173 LNTRLTDDVSKINEGIDKIGMFFQSIATFTFTGTVGTRGKWLTLVLAISPVLGLSAA 232
DB 61 LNTRLTDDVSKINEGIDKIGMFFQMATFTFGFTIGTRGKWLTLVLAISPVLGLSAG 120

QY 233 IWAKILSSFTDELLAYAKAGAVAEVLAARTVIAFGGQKKELEYKNLEEAKGIGIK 292
DB 121 IWAKILSSFTDELLAQYAKAGAVAEVLAARTVIAFGGQKKELEYNNLEEAARLGIK 180

QY 293 KAITANISIGAAFLIYASALAFWYGSVLSSYSISGOVLTVPFVSLIGAFSIGQASP 352
DB 181 KAITANISIGAAFLIYASALAFWYGSVLISKEYISGOVLTVEFVLIAPFISGQASP 240

QY 353 STEAFANARGAAEYFKIIDNKPISIDYSKSGHKPDNKGLEFNKVFHSPSRKEVKIL 412
DB 241 NTEAFANARGAAEYFNIIIDNKPISIDFSKNGYKPDNKGLEFNKIHFSPSRKDVQIL 300

QY 413 KGLNLKVSQGTVALVGNSSGCKSTTVOLMORLYDPTDGMVCDIGQDITRTINVRHLREIT 472
DB 301 KGLNLKVSQGTVALVGNSSGCKSTTVQLQLRLDPTDGVGSIDGQDITRTINVRHLREII 360

QY 473 GYVSQEPVLFAITIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQ 532
DB 361 GYVSQEPVLFAITIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQ 420

QY 533 LSGGQKQRTAIALVRNPKIILLDEATYALDTESEAVVQVALDKARGRRTIVIAHRLS 592
DB 421 LSGGQKQRTAIALVRNPKIILLDEATYALDTESEAVVQVALDKARGRRTIVIAHRLS 480

QY 593 TVRNADVIAGFDGIVVEKGNHDELMKEKGIYFKLVMTQTRGNETELENAESKESDA 652
DB 481 TVRNADIIAGFDGGVIVEGNGHEELMREKGIYFKLVMTQTAGNELELNGEVESKNEIDN 540

QY 653 LEMSPKXSSSLKRRSTRSRTHAPQGDRLKGTCKEDLNENPVVPSFWRLKLNSTWPTY 712
DB 541 LDMSKSDASSLIRSTRSRIRGPHDQDKLSTKEALDEDVPPISFWRLKLNSEWPY 600

QY 713 FVVGIFCAIINGLQAPAFSIIISRIGITFTRDEDPETRONSMESVLEVLVLIISFTIF 772
DB 601 FVVGIFCAIVNGALQAPAFSIIISKVVGVFTRNTDDETRKHSNLSFLLFVLIGVLSFTIF 660

QY 773 FLQGFTFGKAGELTLTKRLRYMFRSLRODVSFDFDPKNTTGALTRLANDAAQVKGAG 832
DB 661 FLQGFTFGKAGELTLTKRLRYMFKSMLRQDVSWFDPKNTTGALTRLANDAGQVKGATG 720
```

```

QY 833 SRLAVITQNIANLGTGIIISLIYGWQLTLLLLAIVPIIAIAGVVMKMLSGQALKDKKEL 892
DB 721 ARLAVITQNIANLGTGIIISLIYGWQLTLLLLAIVPIIAIAGVVMKMLSGQALKDKKEL 780

QY 893 EGAGKIATEAIEINPRVTVTRQKFEYMYAQSLQVYPYRNSLRKAHIFGVSPSITQAMMY 952
DB 781 EGSCKIATEAIEINFRVTVTRQKFEYMYAQSLQVYPYRNSLRKAHIFGVSPSITQAMMY 840

QY 953 FSYAGCFRFGAYLVANEFMNFQDVLVFSVIFGMAVGVSSFPADYAKAKVSAAHVIM 1012
DB 841 FSYAACFRFGAYLVARELMTFENVLLVFSVIFGMAVGVSSFPADYAKAKVSASHIIM 900

QY 1013 IIEKSPILDISYSPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVYKGGQTLALVG 1072
DB 901 IIEKSPISIDSYSTGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLNLEVYKGGQTLALVG 960

QY 1073 SSGCGKSTVVQLLERFYDPLAGSVLIDGKEIKHLNWQLRAHLGIVSQBPILFDCSIAEN 1132
DB 961 SSGCGKSTVVQLLERFYDPMAGTVFLDGKEVNLQNLRAHLGIVSQBPILFDCSIAEN 1020

QY 1133 IAYGDSNRVVSHEEIMQAAKEANIHHFIETLPEKYNTRVGDGKTQLSGGQKQRIATARAL 1192
DB 1021 IAYGDSNRVVSQDEIERAAKEANIHFIESLPKYNTRVGDGKTQLSGGQKQRIATARAL 1080

QY 1193 VRQPHILLDEATYALDTESEKVVQVQALDKAREGRTICIVIAHRLSTIONADLIIVFQNGK 1252
DB 1081 VRQPHILLDEATYALDTESEKVVQVQALDKAREGRTICIVIAHRLSTIONADLIIVFQNGK 1140

QY 1253 VKEHGTHQOQLLAQKGIYFVSMVSVQAGAKR 1281
DB 1141 VKEHGTHQOQLLAQKGIYFVSMVSVQAGAKR 1169

RESULT 7
O93437 PRELIMINARY; PRT; 1288 AA.
ID O93437;
AC O93437;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ABC TRANSPORTER PROTEIN.
GN CMDBI.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RX MEDLINE=99209805; PubMed=10195430;
RA Edelman H.M.L., Duchek P., Rosenthal F.E., Foeger N., Glackin C.,
RA Kane S.E., Kuchler K.;
RT "Cndrl, a chicken P-glycoprotein, confers multidrug resistance and
RT interacts with Estradiol.";
RL Biol. Chem. 380:231-241(1999).
DR EMBL; AJ009799; CAA08835.1; -.
DR HSSP; P13569; INBD.
FT CHAIN 2 1288 ABC TRANSPORTER PROTEIN.
SQ SEQUENCE 1288 AA; 141917 MW; CB258A5F2826DB6C CRC64;

Query Match 72.1%; Score 4665.5; DB 13; Length 1288;
Best Local Similarity 71.3%; Pred. No. 1.1e-260;
Matches 920; Conservative 160; Mismatches 166; Indels 45; Gaps 12;

QY 2 DPEGGRKGSAAKFNWKGKSKKNEKKEKP-TVSTFAMFRYSNWLDRMLVYMGTAAMAI 60
DB 27 DPEDEKG-----KKGKGPQWVSPALFRYSSCTDKLLMIFGSLAIA 71

QY 61 HGAALPLMLVFGNMTDSFANAGISRNKTPFVINESITNTQHTFHNHLEEMTTAYYY 120
DB 72 HGTSLPIAMIIFGDWTDSFTSGMTN-----ITGNSSGLNSADSVFNKLEEMTRYAYY 126
```



```
Db 368 ARGAAVTIFNIIDNOPKIDFSKESGLKPKDKIGDIEFNVIPTSPSKDIOVLKGLNLNI 427
Qy 420 QSGOTVALVGNCGCKSTTVQLMORLYDPTDGMVCDGODIRTNVRLREITGVVSEPP 479
Db 428 PSKKTVALVSGSGCKSTTVQLIRFYDPTDGVITLDGODIRSLNIRYLREITGVVSEPP 487
Qy 480 VLFATTIAENIRYGRNVMTDIEKAVKANNAYDFIMLPNKFDTLVGERGAQLSGGQK 539
Db 488 ILFDTTIANIRYGRDVTKEIERATKANAYDFIMLPDKLETLVGERGTQLSGGQK 547
Qy 540 RTAARALVRNPKILLDDATSEAVVOALDKARKGRTTIVIAHRLSTVRNADV 599
Db 548 RTAARALVRNPKILLDDATSEAVVOALDKARKGRTTIVIAHRLSTVRNADV 607
Qy 600 IAGFDGVIKGNHDELMEKGIYFKLVMTOT---RGNELELENAYSESSEDALEM 655
Db 608 IAGFDGVIKGNHDELMEKGIYFKLVMTOT---RGNELELENAYSESSEDALEM 660
Qy 656 SPKDSGSSILKRRSTRSIIHA--POGQDRKLTKE--DLNENVPVPSFWRILKLNSTW 712
Db 661 PVYTHSNLVRKSSRNTIKSKVPETEDKEVDEEKKKEGPPVPYFKVMKLNKPW 720
Qy 713 FVVGIFCAIINGLOPAFSIISRIIGTIFRDEDPETK--RQNSNMFSLVFLVIGIISFIT 771
Db 721 FVVGIVICAMINGATQPAFAIIFSRIGVFA--GVPVSQMSSESSMYSLFLALGGVSEFIT 777
Qy 772 FFLQGTGFKAGELTKRLRYVWFRSMLRQDYSWFDPPKNTTGALTTRLANDAAOVKGAI 831
Db 778 FFLQGTGFKAGELTKRLRYVWFRSMLRQDYSWFDPPKNTTGALTTRLANDAAOVKGAI 837
Qy 832 GSRILAVITONIANLGTGIIISLYIGQLTLLLAIVPIIAIAGVEMKMLSGQALKDKKE 891
Db 838 GTRLALLAQNANLGTGIIISLYIGQLTLLLAIVPIIAIAGVEMKMLSGQALKDKKE 897
Qy 892 LEGAGIATAEIENFRVTVSLRQEFKFFYAGSLQVYPRNSLRAKHIFGVSEFITQAMM 951
Db 898 LEKAGKISTDAVLNIRTVVSLTRERFEAMYKSELEGPYRNSIKKAHLHGLTYGLSQAH 957
Qy 952 YFSYAGCFR--FGAYLVANEFMNFODVLLVFSATVFGAMAVGVSSFPADYAKAKVSAHV 1010
Db 958 VLCLCWFSVLGAYLVVEGLMKLDEVFLVSSAIVLGAMALGOTSFPADYTRKAMISAAHI 1017
Qy 1011 IMIEKSPIDSYSPHGLKNTLEGWNTFNEVFNTPTRPDIPVLOGLSLEVKKGOTLAL 1070
Db 1018 FSLLRVPQIDSYSDQGEKPKNSGNWFKGVNFNTPTRPDITVLQGLDISVKGQETLAL 1077
Qy 1071 VSGSGCKSTVOLLERFYDPLAGSVLIDGKETKHLNVQWLAHLGIVSQEPTLPCSTA 1130
Db 1078 VSGSGCKSTVOLLERFYDPLAGSVLIDGKETKHLNVQWLAHLGIVSQEPTLPCSTA 1137
Qy 1131 ENTAYGNSRVWSHEETMOAKKANIHHFTETLPEKYNTFRVGDGQTLSGGQKQRIATAR 1190
Db 1138 DNAYGDNRRKVQEELETAKEANIHSFESTDKYNTFRVGDGQTLSGGQKQRIATAR 1197
Qy 1191 ALVRQPHILLDDATSEALDSEKVVQEOALDKAREGTCIVIAHRLSTONADLIWVFON 1250
Db 1198 ALIRKPKILLDDATSEALDSEKVVQEOALDKAREGTCIVIAHRLSTONADLIWVFON 1257
Qy 1251 GKVKHCTHOOLOAKGIVFSWVSQAG 1278
Db 1258 GKVVQEGTHOOLOLKGIVFSVLTIOLG 1285

RESULT 9
Q90235 PRELIMINARY; PRT; 1348 AA.
AC Q90235;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BILE SALT EXPORT PUMP.
OS Raja erinacea (little skate).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
```

```
OC Elasmobranchii; Squala; Hypnosqualea; Pristiorajae; Batoidea;
OX Rajiformes; Rajidae; Raja.
RN NCBI_TaxID=7782;
RP SEQUENCE FROM N.A.
RX MEDLINE=21340059; Pubmed=11447010;
RA Cai S.Y., Wang L., Ballatori N., Boyer J.L.;
RT "Bile salt export pump is highly conserved during vertebrate evolution
and its expression is inhibited by pFIC type II mutations."
PL Am. J. Physiol. Gastrointest. Liver Physiol. 281:G316-G322(2001).
DR EMBL; AF367243; AAK52958.1; --
SQ SEQUENCE 1348 AA; 148648 MW; 819065641654CFDD CRC64;

Query Match 51.5%; Score 3336; DB 13; Length 1348;
Best Local Similarity 49.9%; Pred. No. 6.5e-184;
Matches 662; Conservative 253; Mismatches 354; Indels 58; Gaps 10;

Qy 1 MDPSGGRK-GSAENKFWMGKKSKN-EKEKKKPTVTFAMFRYSNWLDRILMLVGTMAA 58
Db 27 IDGPGQRKVTVNGNYPSPSKSDNENLSKKGRIKIGFFQLFRFASCYEVTLMVFGVCA 86
Qy 59 LIHGAALPLMMLVFGNMTDSFANAGISRNKTFPVIINESITNNTHFIN----- 107
Db 87 LLHGAQPAVLVLVFGLLDTFTFIADIELOELKDT--RKHCNDNTIHWVNGSEYLNQHNET 144
Qy 108 -----HLEEMTYTAYYSIGAGVLVAAYIQVSWCLAAAGROILKIRKOFFHAIMRQE 161
Db 145 MSCGILNIEKEMTFAYYVVGICSVLLGYFCFWTAAAROTQIRKAYFRQIMRME 204
Qy 162 IGWFDVHDVGEINRLDDVSKINEGIGDKIGMEFFQSTATFFTCFIVGTRGWNKLTIVIL 221
Db 205 MGFEDCSVGLNTRMSDDINKINDAIDQVIGIFQRFSTFVSGFLMGFVNGWKLTLVII 264
Qy 222 AISPVGLSAAIWAKILSSFTDKELLAYAKAGAAVEELAAIRTVIAFGGOKKELERYNK 281
Db 265 AVSPLIGLGAALMALSVARLTGLGLKAYAKAGAADEVLSIRTVAASFGEKKEVDYDR 324
Qy 282 NLEAKGIGIKKAITANISGAALLIYASALAFWTGTSVLV--SSEYSIGQVLTVFVS 340
Db 325 NLVFAQRWIRKIMTGFYGVVMIIFCCYALAFWYGSKLIVIQEYTPPGGLQVFLGV 384
Qy 341 LIGAFSIGQASPSTEAFANARAGAAVEIFKIIDNKPSSIDYSKSGHKPDNKGLEFNKH 400
Db 385 LVAAANLQASPCLEAFASGRGAAVKPEITDRPEIDCKSEGGYTLNKYKGDIEFHNVT 444
Qy 401 FSYPSRKEVKILKGLNLKVQSGQTVLVNCGSGCKSTTVQLMORLYDPTDGMVCDIGODI 460
Db 445 ENYPSRPDVKTLLRLSMVIKAGETTAFVPGSGSGKSTAVELIQRFYDPKQGMVTLDGHI 504
Qy 461 RTINVRHLREITGVVSOEPVLFPATTIAENIRYGRNVMTDIEKAVKANNAYDFIMKLPN 520
Db 505 RSLNIQWRLSLIGIVQEPVLFSTIEENIRYVREGLTNDIVQAAKEANAYDFIMDLPE 564
Qy 521 KFTLVGERGAQLSGGQKQRIATARALVRNPKILLDDATSEALDSEAVVQVVALDKARK 580
Db 565 KFTLVGGGGQSGGQKQRIATARALVRNPKILLDDATSEALDSEAVVQVVALDKARK 624
Qy 581 GRTTIVIAHRLSTVRNADVIAAGFDGVIKGNHDELMEKGIYFKLVMTOTRNEIELE 640
Db 625 GRTTIVIAHRLSTVRNADVIAAGFDGVIKGNHDELMEKGIYFKLVMTOTRNEIELE 684
Qy 641 NATGESKESDALEMSPKDSGLIKRSTRSIIHAPOQDRK----- 683
Db 685 KA---RQVNGAIEDGASEK-RQLIRGSSRASVNRSTRHQRSQVSEVLSDLSGPGDVA 739
Qy 684 -----LGTK----EDLNENVPVPSFWRILKLNSTWEPVVFVVGIFCAIINGLOPAF 730
Db 740 SAVRTPSISLGDDEKDVDEEESIEPAPVSRILKYNVSEWYMLFGSLGAANGVNPYI 799
Qy 731 STIFSRIIGTIFRDEDPETKRONSNMFSVLVIGIISFITFFLOGTFFGKAGELLKRL 790
Db 800 ALLFSQILGTFTSL-QNEEEKINQINAICLFFVVGVLVSFLTQFLQSYFFAKSGELLTTRL 858
```


Db 1022 TYPO-LSGEVKNLKVFPYPERPAVPIQGLNVHVKPGQTALVPGSGCKSTVISILLER 1080
QY 1088 FYDPLAGSVLDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAENIAYGDNRSRVSHHEI 1147
Db 1081 LYDPLEGAVTVNDLRLQMPKHLRKHIALVQSPELILFDTISRENIVYGLQPGGYTHEQI 1140
QY 1148 MQAAKEANIIHFTETLPEKYNTRVGDGKGTQSGGQKORIAIARALVRQPHILLDEATSA 1207
Db 1141 ETACKANIHKFIDELPDGTVETRVGEGKGTQSGGQKORIAIARALVFNPKILLDEATSA 1200
QY 1208 LDTESEKVOQALDKAREGRTICIVIAHRLSTIONADLIVVFQNGKVKHEGHTHOOLLAQKG 1267
Db 1201 LDTESEKVOQALDKAREGRTICIVIAHRLSTIONADLIVVFQNGKVKHEGHTHOOLLAQKG 1260
QY 1268 IYFSMVSVOA 1277
Db 1261 AYFALTQKQS 1270

RESULT 12
O01495 PRELIMINARY; PRT; 1265 AA.
ID O01495
AC O01495;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 140.5 KDA PROTEIN.
GN C34G6.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.,"
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Langston Y., Rohlfing T.;
RT "The sequence of C. elegans cosmid C34G6.,"
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.,"
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RW EMBL; U97407; AAB52482.2; --
DR Hypothetical protein.
SQ SEQUENCE 1265 AA; 140465 MW; 4948EF5C5A402757 CRC64;

Query Match 43.78; Score 2832; DB 5; Length 1265;
Best Local Similarity 45.48; Pred. No. 7.3e-155;
Matches 584; Conservative 246; Mismatches 399; Indels 56; Gaps 14;
QY 21 KSKNEK-----KEKKPTVSTFAMFRYSNWLDRMLYLVGTMAAIIHG 62
Db 2 KSRNKPVTWKPLLRKSHSSDSSIDESTVKLTNYGIFYTQGVDELILLITGTVAAVIHG 61
QY 63 AALPLMLVFNQWTDSPANAGISRNKTPPV--IINESNTNQTQPHINLHEEMTYAYYS 121
Db 62 AGFPLLAIVLGMGTTVFLRA---QNSDFVGVGVNVPGLVPSIDFNSFVKKYCYXL 118
QY 122 GIGAGLVAAIYQVFWCLAGROILTKRKOFFHAIMROETGWDFVDHVGELNRLTDDV 181
Db 119 VLGVLVFTTSVQIACFSEYARLVKRLQNYLKAILRQQLQWFDKQOQTGNLTARLTDL 178
QY 182 SKINEGIDKIGMFFQSIATFTTGTFTGTRGWKLTIVLAISPVLGSLAAINAKILSSF 241

Db 179 ERVREGLGDKFALLVQMFAAFLAGYGVGFYSWMTLVMGFAPLIVLSGAMSKSMATR 238
QY 242 TDKELLAYAKACAVAEVLAARTVIAFGGOKKELERNKNLEAKGIGIKKAITANISI 301
Db 239 TRVEQETAVAGAIABETFSIRTVHSLNGHAKREDRFYNALEVGQRGTIVKYVWGLGV 298
QY 302 GAAPLIIYASALAFWYGTSLVSS--EYSIGQVLTVFFSVLIGAFSIOGASPSIPAFANA 360
Db 299 GFSNLCMYSSYALAFWYGSTLIINDPTDFRGLITVFVFAVLSSGSLGALPHLASFGTA 358
QY 361 RGAAYEIFKIIONKPSIDSYSGHKPDNIKLNLEKFNHVSFSPSRKEVKILGLNLAVQ 420
Db 359 RGAASATLVLSVINSHPKIDPYSLGLVDMKGDISFKOVHFRYPKRDHVLKGLGSLDK 418
QY 421 SGTVALVGNSSGCKSTTVQLMORLYDPTDMVCIDGODIRTINVRHLREITGVVSOBPV 480
Db 419 AGDKIALVSSGCKSTIIVNLQRFYDPTKGRVLDGVDLREVNVHSLREQIGIVSQEPV 478
QY 481 LFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPNKEDTLVGERGAOLSGGQKOR 540
Db 479 LFDGTIYENIKMGNEHATHDQVVEACKMANANDFIKRLPDGYGTRVGEKGVOLSGGQKOR 538
QY 541 IAIARALVNPKILLDEATSDLTSEAVQVALDKARKGRTTIVIAHRLSTVNRADVI 600
Db 539 IAIARALVNPKILLDEATSDLTSEAREVQALDQAQAGRTTIIVAHRLSTIRNVRI 598
QY 601 AGFDDGVIVKGNHDELMEKGIYFKLV-----TMOIRGNEIELENATGSKSESAL 653
Db 599 FVKAGNIVSESSHEELSKQGIYDMTQAVVROQQEAGKDIE-----DTISESAHS 652
QY 654 EMSPKDSGSSLIKRRSTRRSIHAPOQQRKLGTKEDLNE--NYPVPSFWNRILKLNSTWMPY 712
Db 653 HLSRKSSTRSAL---SIATSIHQ-----LAEVEECKAPPTSMFKIFKNGDKVGV 700
QY 713 FVVGIFCALINGLOPAFSIIISRIIGITREDDEPKKQNSNMFSVLFLVGLISFTF 772
Db 701 FIGGIFGAFIFGVTVPFALVAFINVSPLAD--QMOANYFVWCGMFLVGLITFFVFG 758
QY 773 FLQGTFTGKAGEILTTRKLRVWFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGATG 832
Db 759 FTSANCLRGCGESLTKMLRFEAFKNLLRQDIAYDRLHGTGKLCRTRFATDAPNVR-VYF 817
QY 833 SRLAVITQNIANLGTGIIISLIYWGQTLTLLLAIVPIAIAGVVEMKMLSGQALDKKEL 892
Db 818 TRLPVVLASIVTICGALGIGFYWGQALILVWVPLVWGGYFEMQMRFGQIRDTQLL 877
QY 893 EGAGKIATEAIENFRVSVLITREQEYMYAQSLQVYPYNSLRKAHIFGVSFSTQAMMY 952
Db 878 EEAGKVASQAVEHIRTVHSLNRQEQPHFTYCEYLREPPFNTNKLKHAHTYGAVFASFOSLIF 937
QY 953 FSYAGCFFRGAYLVANEFMNFQDVLVLSAIVFGAMAVQVSPADYAKAKVSAAHVIM 1012
Db 938 FMTAAAFYLGSIYVNOQAQMPIDIVRVYFAISFCGGMIGNTTSFIPDVVKARLAASLIFY 997
QY 1013 IIEKSPILDSYSPHGL-KPNTLEGNTFNEVFNFRPDPVLPVGLSLEVKKGQTLALV 1071
Db 998 LIEHPPTIDSLSDGIVKPIIT--GNISIRNVFNFRPDKTKVLQGTDLIDKAGKTVALV 1055
QY 1072 GSSGCKSTVQLLREYDFPLAGSLVDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAE 1131
Db 1056 GHSGCKSTIMGLERFYNDQKGMIMDGNITNKLNLSSIREQVCLVSEPTLFDCTIGE 1115
QY 1132 NIAYGNSRVVSHHEEIMQAAKEANIIHFTETLPEKYNTRVGDGKGTQSGGQKORIAIARA 1191
Db 1116 NICYGN-RNVTYQEIIEAAMANIHNFIILGLPDGTDYHVEKGTQSGGQKORIAARA 1174
QY 1192 LVQPHILLDEATSDLTSEKVOQALDKAREGRTICIVIAHRLSTIONADLIVVFQNG 1251
Db 1175 LVRSPSVLLDEATSDLTSEKIVQBALDKAQGRGTCLVIAHRLSTIONSVDVIATVSEG 1234
QY 1252 KVEHGHQQLLAKQGIYFSMVSVO 1276
Db 1252 KVEHGHQQLLAKQGIYFSMVSVO 1276

```
Db 1235 KIVEKGTGHDLEIRKSEIYQKFCETQ 1259
RESULT 13
Q9W693
ID Q9W693 PRELIMINARY; PRT; 851 AA.
AC Q9W693;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE TRANSPORTER HOMOLOG (FRAGMENT).
GN MDR.
OS Fundulus heteroclitus (Killifish) (Mummichog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Fundulidae; Fundulus.
OX NCBI_TaxID=8078;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER, INTESTINE;
RA Cooper P.S., Van Veld P.A., Reece K.S.;
RT "P-glycoprotein related sequences from the mummichog (Fundulus
heteroclitus)".
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF099732; AAD23956.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
FT NON_TER 1
SQ SEQUENCE 851 AA; 93660 MW; 8F2DC4464171420A CRC64;

Query Match 43.6%; Score 2825; DB 13; Length 851;
Best Local Similarity 63.4%; Pred. No. 1e-154;
Matches 541; Conservative 143; Mismatches 153; Indels 16; Gaps 4;

QY 436 STTVQLMQRLYDPTDGMVCIQDQIRTNVRLREITGVVSGEPVLFAITIAENIRYGR 495
DB 1 STTIQLLEFYDPQEGSVSIDGHDIRSLNVRLRGVGVVSGEPVLFAITIAENIRYGR 60
QY 496 NVTMDETEKAVKANAYDFTMKLPNKFDTLVGERGAQLSGGOKORTAIRALVRNPKILL 555
DB 61 DVTEEEKAKKANAYDFIMNLPDRFETLVGDRGTQMSGGOKORTAIRALVRNPKILL 120
QY 556 LDEATSAIDTESAVVQVALDKARKGRITVIAHRLSTVRNADVIAGDDGVIVKGNHD 615
DB 121 LDEATSAIDTESAVVQVALDKARKGRITVIAHRLSTVRNADVIAGDDGVIVKGNHD 180
QY 616 ELMEKKGIFKLVMTQTRGNEIELEENATGESKESDALEMSPKDSG---SSLIKRRSTR 672
DB 181 ELMEKKGIVHTLVMTQTRGNEIELEENATGESKESDALEMSPKDSG---SSLIKRRSTR 235
QY 673 SIHA-----POGDRKLGTDLENVPPVSEFWRILKLNSTEMPYFVVGIFFCALING 725
DB 236 SFAASIGEGKDKQKEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 295
QY 726 LOPAFSIIFRIIGIFTRDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 785
DB 296 IQPLEAVLFESKIITVFA-EPDQTIIRQRANFEFLMVFVGVVGVVGVVGVVGVVGVVGV 354
QY 786 LTRRLRYWFRSMRLQDVSFDDPKNTGALTRLANDAAQVKGAGISRLAVITQNIANL 845
DB 355 LTKLRLGAFKSMRLQDVSFDDPKNTGALTRLANDAAQVKGAGISRLAVITQNIANL 414
QY 846 GTGIITSLIYGVQWLTLLLAIVPIAIAAGVWEMKMLSGQALDKKLEAGKAGIATEAEN 905
DB 4 KGODDERIPILGSSKKSIGEVSKKEEPPITNRGILSLATLTDYVLLAAGTAPCVHGA 63

Db 415 GTGVILAFVYGWELTLLILAVVPVIALAGAVQMKMLTGHAAEDKKLEKAGKIATEAEN 474
QY 906 FRTVSLTREOKFEYMYAQSLOVPYRNSLRKAHIFGVFSITQAMMYFSYAGCFRFGAYL 965
DB 475 IRTVASLIREPKFESLYEENLVVPYKNSOKKAHVYGFTEFSQAMYYFAYACFRFGAYL 534
QY 966 VANEFMNFQDVLVLFSAIVFGAMAVGVSSFPADYAKAKVSAARHIMITEKSLPDISYSP 1025
DB 535 IVEGRMDVEAVFLVTSVAVLFGAMAVEANSFAPNYPKAKMSASHLMMLLNKEPEIDNLSE 594
QY 1026 HGLKPNLTGCVTFNEVVFNYTRDPDIPYLGQLSLSEVKGQTLALVYSGSGCKSTVWOLL 1085
DB 595 RGSPEDFMGVNSFEDYKFNYPSPDVPILRLNLRVKKGETLALVYSGSGCKSTVWOLL 654
QY 1086 ERFYDPLAGSLVLDGKEIKHLMVOMLRAHGLVISOEPILFDCSIAENIAYGDNRSVWSHE 1145
DB 655 ERFYDPRGRVYVMSIDVRLNIHWRISQIGVSEVPVLFDCALAEIAYGDNRSVWSHE 714
QY 1146 EIMQAAKEANTHHFTIETLPEKYNTRVGDGKQTLGSGGOKORTAIRALVRNPKILLDEAT 1205
DB 715 ETEAAAKAANIHNFINELPQKYNTOAGDKGTQSGGOKORTAIRALVRNPKILLDEAT 774
QY 1206 SALDTESEKVVQEALDKAREGRTICIVIAHRLSTVRNADVIAGDDGVIVKGNHD 1265
DB 775 SALDTESEKVVQDADQAKGRTICIVIAHRLSTVRNADVIAGDDGVIVKGNHD 834
QY 1266 KGIFYSVMVSVQAG 1278
DB 835 KGVIHMLVTQLG 847

RESULT 14
Q61301 PRELIMINARY; PRT; 1275 AA.
AC Q61301;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE P-GLYCOPROTEIN.
GN PGP-A.
OS Haemonchus contortus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98226176; PubMed=9566525;
RA Xu M., Molento M., Blackhall W., Ribeiro P., Beech R., Prichard R.;
RT "Ivermectin resistance in nematodes may be caused by alteration of p-
glycoprotein homolog."
RL Mol. Biochem. Parasitol. 91:327-335(1998).
CC EMBL; AF003908; AAC38987.1; -.
DR HSSP; P13569; INED.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1275 AA; 142204 MW; 7439D3836B75787C CRC64;

Query Match 43.6%; Score 2823.5; DB 5; Length 1275;
Best Local Similarity 44.3%; Pred. No. 2.3e-154;
Matches 567; Conservative 273; Mismatches 415; Indels 25; Gaps 12;

QY 8 KGSAEKFNWKGKSKKN---EKKEKKPVSTFAMFRYSNWLDRYMLVGTMAAIFHGA 63
DB 4 KGODDERIPILGSSKKSIGEVSKKEEPPITNRGILSLATLTDYVLLAAGTAPCVHGA 63
```

```
QY 64 ALPLMLVFGNMTDSFANAGISRNKTPFVIINESITNTQTHFHNLEBEMTT---YAY 119
Db : : | | | | | : | | : : : : : : : : : : | | | | | : | |
64 GFSVLGVLGGMTTVFLRA---QNSEF---VLGTVSRDPEGLPALTKKEFDLVRVCLY 117
QY 120 YSGIAGVLVAAIYQVSWFCLAAAGQIILKIRKQFFHAIIMROEIGWFDVHVGELNRLTLD 179
Db : : | | | | | : | | : | | : | | : | | : | | : | | : | | :
118 YLGLGFANFATSYIQIVCWETFAERITHKLKRKYLKAILROQISWFDIQQTGLNLTARLTD 177
QY 180 DVSKINEGIDKIGMFISATFFTFVGTGRGKLTILVLAISPVGLSLAAIWAKLTLS 239
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | : | | :
178 DLVRPEGLGDKLSLFIQVSAFVAGFCVGFAYSWSMTLVMMVAPFIVISANMWSKIVA 237
QY 240 SFTDKELLAYAKAGABAEVLAAIIRTVIAFGGQKKELERYNKNLEEAAGIGIKKAITANI 299
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
238 TRTQVEQETVYAGAAIETESSIRTVHSICGHKRELTFRPEAALEKGRQTLGVKYFYMGV 297
QY 300 SIGAFLIYASYALAFWYGTSLVLS--EYSIGQVVLVFFSVLIGAFSIGQASPSIDAPA 358
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
298 VGFGQMCITYSYALAFWYGSVLINDPALDRGRIFTVFFAVMSGSAALGTCLPLHNTIS 357
QY 359 NARGAAYEIFKIIDNKPSIDYSKSGHKPDNIKGNLEPKNVHFSYPSRKEVKILKGNLKL 418
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
358 IARGAVRSVLVINSRPKIDPSLDGIVLNNMGRSIRKKNVHFSYPSRKTQIILKGVSLQ 417
QY 419 VOSGQTVALVNSGCGSKSTTVQLMORLYDPTDGMVCIDGQDIRTINVRHLREITGVVSQE 478
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
418 VSAGQKIALVSGSGCKSTNVNLLRFYDPTGRKVTIDDDVCDLVQKLRQIIGVVSQE 477
QY 479 PVLFAATTAEINIRVGRNVTWDETEKAVKEANAYDFIMKLPKFDTLVGERGAOLSGGQK 538
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
478 PVLFDGTLFENIKMGYEQATEEVEQACRVANAADFTKRLPEGYGTRVGERGVQLSGGQK 537
QY 539 ORIAIARALVRNPKLILLDEATSDALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNAD 598
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
538 ORIAIARAIINPRILLDEATSDALDTEAESIVQEALKAQGRTTIVIAHRLSTIRNVD 597
QY 599 VIAGFDDGVIVYKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKSDALEMSPK 658
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
598 OIFVFKNCTIVEQTHAELMNKRGVFFEMTQAVLRQKEEVLDSADSDVVSPTALP 657
QY 659 DSGSLIKRRSTRSIHA--PGQDRKLGTKEDLNENVPVSVFWRLKLNSTEWPFVVGVI 717
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
658 HSLSRKSTRSAISAVPSVRSMQI--EMEDLRAKPTMS--KIFYENRDKWGYFTLGL 714
QY 718 FCAIINGLOPAFIIIFSRIGITRDEDPETKRONMF--SVLEFLVLIISFTITFELQ 776
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
715 IACITGTVTFFAVLAQIIQVYS---EPVDQMKGHVFWCGAFIVIGLVHAFAPFSA 771
QY 777 FTFGKAGEILTKRLRYMVFRSMLRQDYSWFDPKNTTGALTTRLANDAAQVKAIGSRLA 836
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
772 ICLGRGGEALPKLRFEEAFKNLLRQNVGFYDDIRHGTVKCLCTREATDPNVR-YVFTRLP 830
QY 837 VITQNIANLGTGIIISLIYGHQLLILLAIIVPITAGVEMKMLSCQALKDKKELEGAG 896
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
831 GVLSSVVTIIGALVIGIFGQALAILMVMVPLIIGSGYFEMRMQFGKKMRDTELLBEAG 890
QY 897 KIATEALENPTVYSLTREORFEMVYAGSLQVVPYRNSLRKAHIFGVSVFSITQAMWYSYA 956
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
891 KVASQAVENRTVALNRQGFHMYCEYLKEPYRENLCQAHYGGVFASFQSLFFMYA 950
QY 957 GCFRFGAYLVANEPFMDQVLLVFSIAIVFGAMAYGVSSFPADYAKAKVAAHVIMTIEK 1016
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
951 VAFWIGAIFVDNHSMQPIDVYRVFFAFMFCGQMGVGNISFPDVKARLAASULLFYLIEH 1010
QY 1017 SPLDYSPLGKLENTLEGNTVNEVNFNTPRPDIPVLOGLSLEVKKGGOTLALVSGSGC 1076
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
1011 PSEIDNLSGDV--TKKISGHISFRNVYFNPTRRQIRVLRGLNLEINPGTFTVALVGSQC 1069
QY 1077 GKSTVOLLERFYDPLAGSVLLDCKEIKHLNVQWLRHGLGVISOEPLIFPCSAENATAYG 1136
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
1070 GKSTVALLERFYNQNKGVITVDGENIRNMNIRNLREQVCIVSOEPLTFDCTINENTCYG 1129
QY 1137 DNSRVWSHEETMQAAKEANIHHTFETLPEKYNTRVGRGKGTQLSGGQKORIAIARALVRQP 1196
```

```
Db 1130 LDOPKPSYEQVAAAKMANIHNFVLGLPEGYDFRVGKGTQLSGGQKORIAIARALIRDP 1189
QY 1197 HILLDEATSDALDTESEKVVQOEALDKAREGRTCVIAHRLSTIONADLIIVFONGKVEH 1256
Db 1190 PILLDEATSDALDTESEKIVQDALEVARQGRTCVIAHRLSTIQDSDVIYMIQEGKATDR 1249
QY 1257 GTHOQLLAQKGIYFSMVSVQ 1276
Db 1250 GTHEHLLMKNDLYKRJCETQ 1269

RESULT 15
Q21349
ID Q21349 PRELIMINARY; PRT; 1321 AA.
AC Q21349;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE K08E7.9 PROTEIN.
GN K08E7.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Snye R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT 'genome sequence of the nematode C.elegans: A platform for
investigating biology.';
RL Science 282:2012-2018(1998).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; 277666; CAB01232.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1321 AA; 145062 MW; 68094121B30E8746 CRC64;

Query Match 42.4%; Score 2746; DB 5; Length 1321;
Best Local Similarity 45.1%; Pred. No. 7.2e-150;
Matches 585; Conservative 221; Mismatches 433; Indels 58; Gaps 12;

QY 21 KSKNKKKKKPTVSTFAMFRYSNWLRLYMLVGTMAAIIHGAALPLMLLVFGNMTDSPA 80
Db : : | | | | | : | | : | | : | | : | | : | | : | | : | | :
47 KITRDAKEEVNKKVSPQLRYRTTLEKLLFLTGTLVAVITGAGLPLMSILQKYSQAFL 106
QY 81 NAGISRNKTPFVIINESITNTQTHF-----NHLEEMTYAYYYSGIAGVLUVAA 132
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
107 NE-----QIVIN----NGSTFLPTGQNTKTDHFDVMMVMSYAAMTVGMAAGQ 154
QY 133 IOVSFCLAAAGROILKIRKQFFHAIIMROEIGWFDVHVGELNRLTDDVSKINEGIDKI 192
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
155 ITVTCYLYVAEQNNKRLREFVKSLIRLQELISWFDTNHSGTLATKLFNLERVEKGDKI 214
QY 193 GMFQSIATFFTGFIYVGTGRGKLTILVLAISPVGLSAAIWAAILSSFTDKELLAYAKA 252
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
215 GMAFVLSQFIITGFIVAFTHSWQLTLMVLAVTPIQALCGFAIAKSMSTAIETLYAKA 274
QY 253 GAVAEVLAARIATVIAFGQKKELERYNKNLEEAAGIGIKKAITANISGAFLIYASY 312
Db : : | | | | | : | | : | | | | : | | : | | : | | : | | :
275 GKVEETISSIRTVVSLNGLRYELERYSTAVEEAKAGVLKGLGISFGMAQSNFISF 334
```

Qy	313	ALAFWYGTSLVLSSEYSIGOVLTVPFVSVLIGAFSIGQASPIEAFANARGAAVEIEPKIID	372
Dy	314	: : : : : : : : : : :	390
Dy	335	ALAFYIGVGWHDGSLNFGDMLTFTSSVMGSMALGLAGPOLAVLGVTAQGAASGIYEVLDD	394
Qy	373	NKPSIDSYSKSGHKKPDKNIKNLEFNKVHFSPSRKEVKYLKGKLNLKYOSGOTVALVCNSG	432
Dy	395	RKPVIDSSSKAGRKMKIKGDITVENVHTTPSPRDPDILRGMLNRVNAGOTVALVGSSG	454
Qy	433	CGKSTTVQLMQRLYDPTDCGMVICODQIRTNVRHLREITGVWSQEPVLFAETIAENIRY	492
Dy	455	CGKSTIISLLRYDYDLKGGKITDGVDYRDINLEFLRNNAVVSQEPALFNCTIEENISL	514
Qy	493	GRENVTMEITEKVAKEANAYDIMTKLPNKFDPLVGERGAQSOGQOKORIAIARALVRNPK	552
Dy	515	KGEKITREENVAACKMANAEKTIKLPNGYNTLDGDRGTQSLGQOQOKORIAIARALVRNPK	574
Qy	553	ILLDDATSALDPESEAVVOVALDKARKGRTTIVIAHRLSTRNADVIATFGDGVIVEKG	612
Dy	575	ILLDDATSALDAESGIVQQALDKAARGRTTIIAHRLSTRNADLIISCKNGVVVEVG	634
Qy	613	NHDELAKKEGIYPFKLVMTROGNEIE-----LENATGESKESDALEMSPKDSGSSL	664
Dy	635	DHRALMAQOGLYDLYTAQTFEDAVDSAAGKFESRENSVARQTSEHGELSRQASEMDIM	694
Qy	665	IK-RRSTRSI-HAQO-----GQDRKLGTKEDLNENVPVFWFRILKLNSTEWPY--	712
Dy	695	NRVRSSTIGSITINGPVIDEKEERIGKDALSRLKOOLEN-----NAQTNLFETILYHA	747
Qy	713	-----FVWGIFCAIINGLGQAPAFSIISRIGIETFREDDETRKQNSNFVSLFLVGI	766
Dy	748	RPHALSUFICGSTATTGGFIYPTYSVFFTSFMNVFA--GNPADFLSQGHFWMFLMLVLA	805
Qy	767	ISFTFFLQGFTEGKAEGILTUKRLRYMVFPSMLRODVSWFDDPNKNTGCALTTRLANDAAQ	826
Dy	806	AQGICSPLMTFFMGIASESLTRDLRNKLFERNVLSQHIGFFDSPONASKISTRLATDPN	865
Qy	827	VKGAIGSR LAVITQNTANLTGTLIIISLYGWOLTLTLLAIIVPIITAAGVEMKMISOAL	886
Dy	866	LRTAIDRFSTVITLIVSWGAGIFFGWMALLIITAILPIVAFQGYLRGRRTGKNV	925
Qy	887	KDKKELEGAGKIATEIENFRVWSLITREQFEYMYAQSLQVPYRNSLRKAHIFGVFSFI	946
Dy	926	KSASEFADSGKIAIEIENVRTVQALAREDTYENCFEKLDIPHEKAIKEAFIQGLSYGC	985
Qy	947	TQAMWFYSYACGRFGAYLVANEFNFDQ--VLLVFSYAIVFGAMAGOVSSFAPDIYAKAK	1004
Dy	986	ASSVLYLLNTCAYRMGLALIITDPPTMPMRVRYMYAITISTLTGFAFYSPFPEYAKAT	1045
Qy	1005	VSAHVMIITEKSPLDISPHCLKPNTLEGNVNTENEVVENPRPDIPVLOGLSLBVKK	1064
Dy	1046	PAGGIIFGMRUKISKIDSLSLAGEK-KLYGKVIKFNVRPAYPERPETEILKGLSFSEVP	1104
Qy	1065	GOTLALVGSSGCKSTVWOLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHGLIVISQBPIIL	1124
Dy	1105	GOTLALVGPSCGCKSTWALLERFYDTLGGFEIFDGSSEIKTLNPETHRSQIAIVSQBP TL	1164
Qy	1125	FDCSIAENIAYGNSRWVSHEETMQAKEANTHHFIETLPKYWTRVGDGKTQLSGGQOK	1184
Dy	1165	FDCSIAENIITYGLDPSVYMAQVEEAARLANTHNFIAELPEGFTFTRVGDGTQLSGGQOK	1224
Qy	1185	RIATARALVROPHILLDDATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTIQNALD	1244
Dy	1225	RIATARALVRNPKILLDDATSALDTESEKVVQEALDRAREGRTCIVIAHRLNVTMNADC	1284
Qy	1245	IWVFQNKVKHEGTHOOLLAQKGIYPSVMSVVOQAAGR	1281
Dy	1285	IADVNSGTITTEKHTHTQLMSEKYGKYYKLTQKQTEKK	1321

GenCore version 5.1.13
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 19:26:40 ; Search time 2230.96 seconds
(without alignments)
7749.846 Million cell updates/sec

Title: US-09-672-725c-23

Perfect score: 6474

Sequence: 1 MDPGGRKGSAENFWKMGK.....LLAQKGIYFMSVQAGAKR 1281

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US09672725/runat_04112002_124342_18001/app_query.fasta_1.6435
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725.ecgn.1.1.7365 @runat_04112002_124342_18001 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1421	21.9	943	9 AL520322	AL520322 AL520322
2	1073.5	16.6	981	10 BF969667	BF969667 602272046

	3	1035	16.0	894	10	BF584668	BF584668
	4	1034	16.0	803	10	BG298756	BG298756
	5	1024	15.8	1019	10	BG248052	BG248052
	6	968.5	15.0	944	10	BF796582	BF796582
	7	958	14.8	851	10	BF969062	BF969062
	8	950	14.7	785	9	AV709991	AV709991
	9	916.5	14.2	795	10	BF313560	BF313560
	10	916	14.1	726	10	BG293345	BG293345
	11	910	14.1	587	10	BM053000	BM053000
	12	908.5	14.0	628	10	BG080311	BG080311
	13	889	13.7	563	10	BI964198	BI964198
	14	889	13.7	563	10	BI964218	BI964218
	15	889	13.7	563	10	BM052878	BM052878
	16	880	13.6	605	9	AV616675	AV616675
	17	875	13.5	563	10	BM053038	BM053038
	18	875	13.5	2876	11	AK014319	AK014319
	19	826	12.8	546	10	BI964137	BI964137
	20	821.5	12.7	1201	10	BF304387	BF304387
	21	806.5	12.5	871	12	A2682350	A2682350
	22	805	12.4	759	10	BE376459	BE376459
	23	792	12.2	662	10	BG963638	BG963638
	24	791.5	12.2	886	12	BI139685	BI139685
	25	780	12.0	517	10	BI964115	BI964115
	26	778.5	12.0	897	12	A2541090	A2541090
	27	772	11.9	780	10	BG587938	BG587938
	28	768	11.9	822	10	BI332761	BI332761
	29	766.5	11.8	932	12	A2670821	A2670821
	30	765	11.8	560	10	BE749379	BE749379
	31	764.5	11.8	834	12	A2548312	A2548312
	32	762.5	11.8	947	12	A2683753	A2683753
	33	761.5	11.8	913	12	BM013981	BM013981
	34	760	11.7	669	10	BM013981	BM013981
	35	758.5	11.7	880	12	A2687805	A2687805
	36	750.5	11.6	823	12	A2532602	A2532602
	37	748.5	11.6	853	12	A2679807	A2679807
	38	748.5	11.6	919	12	A2690701	A2690701
	39	745	11.5	699	9	BB667773	BB667773
	40	744.5	11.5	511	9	AA239727	AA239727
	41	743.5	11.5	886	12	A2540627	A2540627
	42	743	11.5	600	10	BM486593	BM486593
	43	741.5	11.5	891	12	A2682250	A2682250
	44	734	11.3	760	10	BM016204	BM016204
	45	730.5	11.3	899	12	BH154857	BH154857

ALIGNMENTS

RESULT 1
AL520322
LOCUS AL520322 LTI_NFL004_NBC2 Homo sapiens cdna clone CS0DB006YCL15 5
DEFINITION prime, mRNA sequence.
ACCESSION AL520322
VERSION AL520322
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 943)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..943
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB006YCL15"
/clone_lib="LTI_NFL004_NBC2"


```
Qy 813 ThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGly 832
|||||
Db 61 ACTGGAGCATTTGACTACAGGCTGCCAATGATGCTCTCAAGTTAAAGGGCTATAGT 120
Qy 833 SerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSer 852
Db 121 TCCAGGCTTGCTGTAATACCAGAAATATACCAATCTTGGACAGCAATAATATATATCC 180
Qy 853 LeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuLeuAlaIleValProIleIleAlaIle 872
|||||
Db 181 TTCATCTATGTTGGCAACTAACACTGTTACTCTTAGCAATTTGACCATCATTTGCAATA 240
Qy 873 AlaGlyValValGluMetLysMetLysSerGlyGlnAlaLeuLysAspLysLysGluLeu 892
Db 241 GCAGGAGTTGTGAATGAATAATGTTGCTGGCAACGACTGAAAGATAGAAAGAACTA 300
Qy 893 GluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeu 912
Db 301 GAAGGTGCTGGGAAGATCGCTACTGAACCAATACAAACATCCGACCGTTGTTCTTTG 360
Qy 913 ThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsn 932
|||||
Db 361 ACTCAGGAGCAGAAAGTTTGAACATATGATGCTCAGAGTTTGCAGGTACCATACAGAAAC 420
Qy 933 SerLeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyr 952
Db 421 TCTTTGAGGAAGACACACATCTTTGGAAATTTACATTTTCCCTTACCCAGGCAATGATGAT 480
Qy 953 PheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsn 972
Db 481 TTTTCCATGCTGGATGTTCCGGTTTGGAGCCTACTTGTGGCACATAAATCATCAGC 540
Qy 973 PheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGln 992
|||||
Db 541 TTTGAGGATGTTCTGTAGTATTTTCAGCTGTTGCTGTTGCTGCGCATGGCGGCGCA 600
Qy 993 ValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMet 1012
Db 601 GTCAGTTTCATTTGCTCTGCTACTATGC-AAAGCCAAATATACAGCAGC-CACATCATG 658
Qy 1013 IleIleGluLysSerProLeuIle---AspSerTyrSerProHisGlyLeuLysProAsn 1031
Db 659 ATCATTTGAAAAACCCCTGTTGATTTGACAGTTGACAGCACGGGAGGCTTAATTTGGCGGA 718
Qy 1032 ThrLeu---GluGlyAsnVal---ThrPheAsnGluValValPheAsnTyrProThrArg 1049
|||||
Db 719 ACACTTTGGGAGGGAATGCTCCTCTTGTGTGGAGTTGGATTTCAATGATTCACCCCGA 778
Qy 1050 ProAspIleProValLeuGlnGly-----LeuSerLeuGluValLysGlyGlnThr 1067
|||||
Db 779 CCGGAATTCACGATGTTTCAAGGATGAGCCCGGGGTAAAGGGCCAAACGGGCTCTCTTT 838
Qy 1068 LeuAlaLeuValGly-----SerSerGly 1075
|||||
Db 839 GGAGCATGTGTGGGAAACCATGTTGCTCTGGG 871

RESULT 3
BF584668 894 bp mRNA linear EST 12-DEC-2000
LOCUS 602098406f1 NCI_CGAP_Co24 Mus musculus cdna clone IMAGE:4218385 5',
DEFINITION mRNA sequence.
ACCESSION BF584668
VERSION BF584668.1 GI:11658386
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 894)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
```

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Place: LLAM9798 row: 1 column: 02
High quality sequence stop: 651.

FEATURES

Location/Qualifiers

source

1..894
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="IMAGE:4218385"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 263 a 203 c 242 g 185 t

ORIGIN

Alignment Scores: 9,34e-86 Length: 894
Pred. No.: 1035.00 Matches: 225
Score: 92.48% Conservative: 21
Percent Similarity: 84.59% Mismatches: 15
Best Local Similarity: 15.99% Indels: 9
Query Match: 10 Gaps: 1
DB:

US-09-672-725c-23 (1-1281) x BF584668 (1-894)

Qy 335 ThrValPhePheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIle 354

Db 2 ACTGTCTTCTTTCCGTGTTAAT-GGAGCATTCAGTTGTTGGACAGCAITCCAAATAT 60

Qy 355 GluAlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLys 374

Db 61 GAAGCCTTCGCCAATCACGAGGAGCAGCTTATGAAGTCTCAAAATAATTTGATAATAAG 120

Qy 375 ProSerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeu 394

Db 121 CCAGATATAGACGCTTCTCAAGAGTGGGCACAAACACACACATACAGAAATCTG 180

Qy 395 GluPheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeuLysGly 414

Db 181 GAATTTAAGATATTCACITTCACCTACCCATCTCGAAAGAGAGTTTCAGATCTTGAAGGCG 240

Qy 415 LeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGly 434

Db 241 CTCAATCTGAAGGTGAAGAGGACAGCGTGGCCCTGTTGGCAACAGTGGCTGTGA 300

Qy 435 LysSerThrThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCys 454

Db 301 AAAAGCACAACTGTCCAGCTGTGCAAGAGGCTCTACAGCCCCCTAGCATGGTCAGT 360

Qy 455 IleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThrGlyVal 474

Db 361 ATCGCGGACAGGACATCAGAACCAATCATGTGAGTATCTGAGGAGCATCATTTGTTG 420

Qy 475 ValSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArg 494

Db 421 GTGAGTCAGGAACCTGTGCTGTTTGCACCCACGATCGCGAGAACATTCGCTATGCCGA 480

Qy 495 GluAsnValThrMetAspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 514

Db 481 GAAGATGTCACCATGATGATGATTTGAAAGCTGTCAAGGAAGCCCAATGCCATGATCTTC 540

Qy 515 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSer 534


```
|||||
Db 541 ATCATGAACTGCCACCAATTTGACACCTGGTGGTGAGAGGGCGGACGTGAGT 600
Qy 535 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu 554
Db 601 GGGGAGAGAAACAGAGAATCGC-ATTGCCGGGGCCCTGGTCGGCAATCCCAAGTGCCCT 659
Qy 555 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValGlnValAla 574
Db 660 TC-GTGGAGCAAGCCACCTCAGC-CGTGGATACAGAAGTGAAGCTTCAGGCCGCA 717
Qy 575 -LeuAspLysAlaArgLysGlyArgThrIle-----ValIleAlaHisAr 590
Db 718 CCTAGTAGGGTACAAAGGCGNGGACCATTTGGGAAGCTCATCGCTGTTACCGTCGAA 777
Qy 590 GluSerThrValArg 595
Db 778 GCTGACGGCCATCGG 793

RESULT 4
BG298756 803 bp mRNA linear EST 21-FEB-2001
LOCUS 602396681F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4511157 5',
DEFINITION mRNA sequence.
ACCESSION BG298756
VERSION BG298756.1 GI:13063728
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 803)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
Clone Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: LLAM10394 row: b column: 22
High quality sequence stop: 698.

FEATURES
source
1. 803
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4511157"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 240 a 162 c 205 g 196 t
ORIGIN

Alignment Scores:
Pred. No.: 1.01e-85 Length: 803
Score: 1034.00 Matches: 223
Percent Similarity: 91.51% Conservative: 14
Best Local Similarity: 86.10% Mismatches: 21
Query Match: 15.97% Indels: 4
DB: 10 Gaps: 0

US-09-672-725C-23 (1-1281) x BG298756 (1-803)
Qy 236 LysIleLeuSerSerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaVal 255
```

```
|||||
Db 3 AAAGATATGTCTTCTTACTGTAAAGAACTCCATGCTTATGCAAAAGCTGGAGCAGTT 62
Qy 256 AlaGluGluValLeuAlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGlu 275
Db 63 GCTGAAGAAGTCTTAGCAGCCATCAGAACTGTGATGCGTT-GGAGGACAAAAGAGAA 121
Qy 276 LeuGluArgTyrAsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIle 295
Db 122 CTTGAAAGGTACAATAACAACCTTGAGAGAGCTAAAAGCTGGGGATAAAGAAAGCTATC 181
Qy 296 ThrAlaAsnIleSerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAla 315
Db 182 ACGGCCAACATCTCCATGGGTGCAGCTTTCTCCTTATGTCATCATATGCTCTGGCA 241
Qy 316 PheTrpTyrGlyThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThr 335
Db 242 TTCGGTATGGGACTTCCTTGGTCATCTCCAAAGAATACTCTAT-GGACAAGTCTCACT 300
Qy 336 ValPhePheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGlu 355
Db 301 GTCCTCTTTCCGTGTTAAATGGAGCATTCAGTCTTGACAGGCATCTCCAAATATTGAA 360
Qy 356 AlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysPro 375
Db 361 GCCTTCGCCAATGACAGGAGCAGCTTATGAAGTCTTCAAAATAATTGATAATAAGCCC 420
Qy 376 SerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGlu 395
Db 421 AGTATAGACAGCTTCTCAAGAGTGGGCACAAACACGACACATACAAGGAATCTGGAA 480
Qy 396 PheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeu 415
Db 481 TTTAAGATATATTCATCTTCAGTTACCCATCTCGAAAAGAAAGTTTCAGATCTTGAAGGCGCTC 540
Qy 416 AsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLys 435
Db 541 AATCTGAAGGTGAAGACGGACAGCGGTGGCCCTGTGGTGGCAACAGTGGCTGGGAAA 600
Qy 436 SerThrThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIle 455
Db 601 AGCACAACTGCCAGCTGATGTCAAAGGCTTACGACCCCTA-GATGGCATGGTCAGTATC 659
Qy 456 AspGlyGlnAspIleArgThrIle-AsnValArgHisLeuArgGluIleThrGlyValva 475
Db 660 CACGGACAGGACATCAGAACCATCCAAATGTGAGGTATCTGAGGGGATCATGGGGTGGTG 719
Qy 475 LserGlnGluProValLeuPheAlaThrIleAlaGluAsnIleArgTyrGly 493
Db 720 AGTCAAGGAACCTGGCTGTTGTTGCCAACCCAGATCGCGAGAAAATTCGTTATGGG 774

RESULT 5
BG248052 1019 bp mRNA linear EST 13-FEB-2001
LOCUS 602359987F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4488404 5',
DEFINITION mRNA sequence.
ACCESSION BG248052
VERSION BG248052.1 GI:12757867
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1019)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLAM10334 row: n column: 21
High quality sequence stop: 650.
Location/Qualifiers

1. .1019
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4488404"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sali; Site_2: Noti; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 276 a 262 c 283 g 196 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 1.16e-84 Length: 1019
Score: 1024.00 Matches: 216
Percent Similarity: 90.48% Conservative: 12
Best Local Similarity: 85.71% Mismatches: 19
Query Match: 15.82% Indels: 5
DB: 10 Gaps: 2

US-09-672-725c-23 (1-1281) x BG248052 (1-1019)

Qy	1033	LeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIle	1052
Db	1	TTGAGAGAAATGCGAATTTAGTGAGAGTCGTGTCAACTATCCACCCGACCCAGCATC	60
Qy	1053	ProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGly	1072
Db	61	CCAGTGTTCAGGGCTGAGCCTTGAGTGAAGAAGGCCAGAGCGTGGCCCTGGTGGC	120
Qy	1073	SerSerGlyCysGlyLysSerThrValValGlnLeuGluArgPheTyrAspProLeu	1092
Db	121	AGCAGTGGCTCGGGAAGAGACACAGTGTCCAGTGTCCAGCGCTTCTACGACCCCATG	180
Qy	1093	AlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArg	1112
Db	181	GCTGGATCAGTGTTCAGATGGCAAGAAATAAAGCACTGAATGTCCAGTGGCTCCGA	240
Qy	1113	AlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsn	1132
Db	241	GCACAGCTGGCATTGTGCCAGAGCCCATTTCTTTGACTGCACATCGCAGAGAAC	300
Qy	1133	IleAlaTyrGlyAspAsnSerArgValValSerHisGluIleMetGlnAlaAlaLys	1152
Db	301	ATTGCTACGGAGACACAGCGGGTGTCTTATGAGGAGATTGTGAGGGCAGCCAAAG	360
Qy	1153	GluAlaAsnIleHisHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGly	1172
Db	361	GAGGCCAACATCCACCACTGATCGATCGCTACCTGATAAATACAACACAGAGTAGGA	420
Qy	1173	AspLysGlyThrGlnLeuSerGlyGlyGlnArgIleAlaIleAlaArgAlaLeu	1192
Db	421	GACAAAGGCACTCAGCTGTGCGGTGGCAGACAGCGCATCGCCATCGCAGCGCCCTC	480
Qy	1193	ValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer	1212
Db	481	GTCAGACAGCCTACATTTTACTTCTGGAGGAACAACATCAGCTCTGGATACAGAAAGT	540
Qy	1213	GluLysValValGlnGlnAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIle	1232
Db	541	GAAGAAGTGTGCCAGGAAGCGCTGGCAAAAGCCAGGGAAGCGCCACCTCATTTGTATC	600
Qy	1233	AlaHisArgLeuSerThrIleGlnAsnAlaAspLeuIle-ValVal-PheGlnAsnGlyL	1252

Db	601	GCTCACGCTGTCCACCATCCAGACGGGACTTGATCGGTGATTCAAGAACC	660
Qy	1252	ysValLys---GluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheS	1271
Db	661	AGGTCAAGGAGCAGCGGACCCCAACGCAAGTGTGGCGCAGAAAGGCATCTACTTCT	720
Qy	1271	erMetValSer-ValGlnAlaGlyAlaLys	1280
Db	721	CCAGG---TCAGGTGCATGCTGGCCCAAG	747
RESULT 6			
BF796582			
LOCUS	BF796582	944 bp	mRNA linear EST 12-JAN-2001
DEFINITION	602258463F1 NIH_MGC_85	Homo sapiens	CDNA clone IMAGE:4341710 5',
ACCESSION	BF796582		mRNA sequence.
VERSION	BF796582.1	GI:12101636	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 944)		
AUTHORS	NIH-MGC	http://img.nci.nih.gov/ .	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: Louis Staudt, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
FEATURES	plate: LLAM9955 row: n column: 15 High quality sequence start: 9 High quality sequence stop: 669. Location/Qualifiers		
source	1. .944		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4341710"		
	/clone_lib="NIH_MGC_85"		
	/tissue_type="lymphoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: Noti; Site_2: Sali; Cloned unidirectionally; oligo-dr primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."		
BASE COUNT	265 a 241 c 237 g 201 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.51e-79	Length:	944
Score:	988.50	Matches:	210
Percent Similarity:	83.64%	Conservative:	20
Best Local Similarity:	76.36%	Mismatches:	39
Query Match:	14.96%	Indels:	8
DB:	10	Gaps:	1
US-09-672-725c-23 (1-1281) x BF796582 (1-944)			
Qy	988	MetAlaValGlyClnValSerSerPheAlaPro-Asp-TyrAlaLysAlaLysValSerA	1007
Db	1	GTGGCTCTAGGACATCCAGTTCATTTGCTCCAGACGATATGCTAAAGCTAGCTCTG	60
Qy	1007	laAlaHisValIleMetIleIleGluLysSerProLeuIleAspSerTyrSerProHisG	1027
Db	61	CAGCCCACTTATTCATGCTGTTTGAAGAAGAACCTCTGTATTGACAGCTACAGTGAAGAG	120

```
QY 1027 lyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrP 1047
Db 121 GGCTGAAGCCTCAAAATTTGAAGGAAATAATAACATTTAATCAAGTCGTGTCAACTATC 180
QY 1047 roThrArgProAspLeuProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnT 1067
Db 181 CCACCGAGCAACCTGGCAGCTGCTTCAAGGGCTGAGCTGGAGTGGAAGAGCCAGCA 240
QY 1067 hrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuGluA 1087
Db 241 CACTAGCCCTGGTGGCAGCAGTGGCTGTGGGAGAGACCGGTGCTCCAGCTCTGGAGC 300
QY 1087 rgPheTyrAspProLeuAlaGlySerValLeuLeuAspGlyLysGluLeuLysHisLeuA 1107
Db 301 GGTCTACGACCCCTTGGCGGGACAGTCTTCGATGGTCAAGAACCAAGAACTCA 360
QY 1107 snValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspC 1127
Db 361 ATGTCCAGTGGCTCAGAGCTCAACTCGGAATCGTCTCTCAGSAGCCTATCCTATTGACT 420
QY 1127 ysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgValValSerHisGluGluI 1147
Db 421 GCAGCATTCGCCAGAAATATGGCTATGGAGACACACGCCGGTGTATACAGAGTGA 480
QY 1147 leMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIleGluThrLeuProGluLysT 1167
Db 481 TCGTGAGTCAGCCCAAGCTGCCAATACATACATCTTTTCATCGAGACGTTACCCCAAAAT 540
QY 1167 yrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleA 1187
Db 541 ATGAACAAGATGGGAGATAAGGGGACTCAGCTCTCAGGAGGTGCACAAACAGAGGATG 600
QY 1187 laIleAlaArgAlaLeuValArgGln-ProHisIleLeuLeuLeuAspGluAlaThrSer 1206
Db 601 CTAATGCCCGACCCCTCATCAGACAAACCCCTCAATCTCTCTGGATGAGATGACATCA 660
QY 1207 Ala-LeuAspThr-----GluSerGluLysValValGlnGluAlaLeuAspLysAlaAr 1224
Db 661 GCCTCTGGATACCTGACAAAGTGAACACAGGTTGTCCCAAGAGCCCTCTGCACACAGCAG 720
QY 1224 gGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAlaAspLe 1244
Db 721 AGAGCGCCGCACCTGCTGTGATTCGTCACCGCTGTC--AACATCCAGAATGCGAGACTT 778
QY 1244 uileValValPheGlnAsnGlyLysValLysGluHis 1256
Db 779 -ATAGTGGTGTCCCAACGGGAGAGTCAAGACACAT 814

RESULT 7
BF969062
LOCUS
DEFINITION
60226993F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4358135 5',
mRNA sequence.
ACCESSION
BF969062
VERSION
BF969062.1 GI:12336277
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 851)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs+email.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

```
Plate: L1AM9995 row: j column: 24
High quality sequence stop: 612.
FEATURES
source
1..851
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4358135"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT 243 a 172 c 214 g 222 t
ORIGIN
US-09-672-725c-23 (1-1281) x BF969062 (1-851)
QY 826 GlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeu 845
Db 3 CAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCATGTAATACCCAGAAATATAGCAATCTT 62
QY 846 GlyThrGlyIleIleLeuSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAla 865
Db 63 GGGACAGGAATATATATCTTCATCTATGTTGGCACTAACACTGTTACTCTTAGCA 122
QY 866 IleValProIleIleAlaIleAlaGlyValValGluMetLysMetLeuSerGlyGlnAla 885
Db 123 ATTGATCCCATCATTCGAATAGCAGGAGTTGTTGAATGAAATGTTGCTGGACAAGCA 182
QY 886 LeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsn 905
Db 183 CTGAAGATAGAAGAAGCACTAGAGGTCCTGGGAAGATCCGCTACTGAGCAATAGAAAAC 242
QY 906 PheArgThrValValSerLeuThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSer 925
Db 243 TTCCGAACCGTGTCTTCTTTGACTCAGGAGCAGAGTTTGAACATATGATGCTCAGAGT 302
QY 926 LeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGlyValSerPheSer 945
Db 303 TTGCAGGTACCATACAGAAACTCTTTGAGGAAGACACATCTTTGGAATATGATTTTCC 362
QY 946 IleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeu 965
Db 363 TTCAACCCAGGCAATGATGATTTTCTCTATGCTGGATGTTCCGGTTTGGAGCTACTTG 422
QY 966 ValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSerAlaIleValPhe 985
Db 423 GTGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTATTATTTTCAGCTGTGTCTTT 482
QY 986 GlyAlaMet-AlaValGlyGlnValSerSerPheAlaProAspTyrAlaLysAlaLysVa 1005
Db 483 GGTGCCATGGCCCGTGGGCAAGTCAGTTCTATGCTGCTGACTATGCCAAAGCCAAAT 542
QY 1005 lSerAlaAlaHisValIleMetIleLeuLysSerProLeuIleAspSerTyrSerPr 1025
Db 543 ATCAGACGCCCATCATCATCATGATCATTTGAAAAACCCCTTTGAT-GACAGCTACAGCAC 601
QY 1025 oHisGlyLysProAsnThr---LeuGluGlyAsnValThrPhe-----AsnGluVa 1042
Db 602 GGAAGGCCTAATTCGCCGAACACATATTGGGAAGGAAATGTCACCATGTGTGTGAACCTGT 661
QY 1042 lValPheAsnTyrProThrArg-ProAspIleProValLeuGlnGlyLeuSerLeuGluV 1062
```

```

Db 662 TTTTCAAAATTTTCAACCCGTAGCGGACATCCAGTCTCCAGGACTGAAGCTGGGGT 721
QY 1062 aLLysLysGlyGlnThrLeuAlaLeu-----ValGlySerSer----- 1074
Db 722 GAAGAAAGGGCCAAAAGCGGTCTCGGGTGGCCACCAATGGTGTGGGAAAGACACCA 781
QY 1075 -----GlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrA 1090
Db 782 GGGGGCCAGTCCGGGAACCGGTTCTAAACC----- 813
QY 1090 spProLeuAlaGlySerValLeuLeuAspGlyLysGlu 1102
Db 814 -----CTGGCGGAAATGCTGCTCTGGGCAAAAAAG 846

RESULT 8
AV709991 785 bp mRNA linear EST 09-OCT-2000
DEFINITION AV709991 Cu Homo sapiens cDNA clone CUAJA09 5', mRNA sequence.
ACCESSION AV709991
VERSION AV709991.1 GI:10728272
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 785)
AUTHORS Peng, X., Song, H., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N.,
Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu
W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J.,
Chen, Z. and Han, Z.
Homo sapiens cDNA Cu clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
Location/Qualifiers
1..785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Cu"
/tissue_type="adrenal cortico adenoma for Cushing's
syndrome"
/dev_stage="Adult"
/note="Vector: pBluescript sk(-)"
BASE COUNT 210 a 180 c 211 g 183 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 6.2e-78 Length: 785
Score: 950.00 Matches: 208
Percent Similarity: 86.97% Conservative: 17
Best Local Similarity: 80.31% Mismatches: 33
Query Match: 14.67% Indels: 4
DB: 9 Gaps: 0

US-09-672-725C-23 (1-1281) x AV709991 (1-785)
QY 955 TyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGln 974
Db 13 TACTAAGGCGCTCCAGCCCTTTGGCAATTCACAGGATATGTTTACTTCCCTTGATAAGAGA 72
QY 975 AspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSer 994
Db 73 GGAATTGGAAGAGTATTTTCAGCTGTGCTTTGTCATGGCGGCGTGGGCAAGTCAGT 132
QY 995 SerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIle 1014

```

```

Db 133 TCATTTCCTCTGACTATGCCAAGCCAAAATATCAGCAGCCACATCATCATGATCAT 192
QY 1015 GluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGlu 1034
Db 193 GAAAAAACCCCTTTGATTGACAGCTACAGCAGCGAAGGCGCTAATGCCGAACACATTTGAA 252
QY 1035 GlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIleProVal 1054
Db 253 GGAATGTGCATTTGTTGAGTTGTTAATCACTATCCACCCAGCGGACATCCAGTG 312
QY 1055 LeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSer 1074
Db 313 CTTCAGGAGCTGAGCCTGGAGTGAAGAGGCGCAGAGCTGGCTCTGTGGTGGCAGCAGT 372
QY 1075 GlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGly 1094
Db 373 GGCTGTGGGAAGAGCAGAGTGGTCCAGCTCTGGAGCGGTCTACGACCCCTTTGGCAGGG 432
QY 1095 SerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHis 1114
Db 433 AAGTGTCTGTGATGGCAAGAAATTAAGCGACTGAATGTTTCAGTGGCTCCGAGCACAC 492
QY 1115 LeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAla 1134
Db 493 CTGGGCATCGTGTCCAGGAGCCATCTCTGTTGACTGCGAGCATGCTGAGAACATTTGCC 552
QY 1135 TyrGlyAspAsnSerArg-ValValSerHisGluGluIleMetGlnAlaLysGluAl 1154
Db 553 TATGGAGACACAGCCGGGTGGTGTACACAGAGAGATTGTGAGGCGCAGCAAGGAGGC 612
QY 1154 aAsnIleHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLy 1174
Db 613 CAACATACATGCTTCATCAGTCACTGCTAT-AAATATAGCATAAAGTANGAGACAA 671
QY 1174 sGlyThrGlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValAr 1194
Db 672 AGGACCTCAGCTCTCTGGGGCCAGAAACA-CACATTGCCATAGCTTGCCCTGGT-AG 729
QY 1194 gGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer 1212
Db 730 ACAGACTATATAATTTGTTTTTGGATGAAGCCAGTCAGCTCTTTGATACAGAAAGT 784

RESULT 9
BF313560 795 bp mRNA linear EST 21-NOV-2000
LOCUS 601900192F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4129222 5',
DEFINITION mRNA sequence.
ACCESSION BF313560
VERSION BF313560.1 GI:11261583
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 795)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@phs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC1023 row: h column: 23
High quality sequence stop: 674.
FEATURES
Location/Qualifiers
1..795
/organism="Homo sapiens"
/db_xref="taxon:9606"
source

```

```
/clone="IMAGE:4129222"  
/clone_lib="NIH_MGC_19"  
/tissue_type="neuroblastoma"  
/note="Organ: brain; Vector: pOTB7; Site_1: xhoI; Site_2:  
EcORI; cDNA made by oligo-dT priming. Directionally  
cloned into EcORI/XhoI sites using the following 5'  
adaptor: GCACGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH_MGC Library."  
BASE COUNT 213 a 189 c 236 g 157 t  
ORIGIN
```

```
Alignment Scores:  
Pred. No.: 8.15e-75 Length: 795  
Score: 916.50 Matches: 210  
Percent Similarity: 87.50% Conservative: 14  
Best Local Similarity: 82.03% Mismatches: 29  
Query Match: 14.18% Indels: 9  
DB: 10 Gaps: 2
```

US-09-672-725C-23 (1-1281) x BF313560 (1-795)

```
QY 1017 SerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsn 1036  
Db 2 AACCCCTTTGATGACAGCTACAGCAGCGAAGCCCTAATCCCGAACACATG-GAAGGAAT 60  
QY 1037 ValThrPheAsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGln 1056  
Db 61 GTCACATT-GGTGAAGT-GTATTCAACTATCCACCGACCGACATCCAGTGTTCAG 118  
QY 1057 GlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSerglyCys 1076  
Db 119 GGACTGAGCTGGAGGTGAAGAAGGCGCAGCTGCTGCTGGTGGCAGCAGTGGCTGT 178  
QY 1077 GlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerVal 1096  
Db 179 GGGAAAGACACAGTGTGTCAGCTCCGAGCGGTCTACGACCCCTTGGCAGGGAAGTG 238  
QY 1097 LeuIleAspGlyLysGluLeuLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGly 1116  
Db 239 CTGCTTGTATGGCAAGAAATAAAGCGACTGAATGTTCACTGGCTCCGAGCACACCTGGGC 298  
QY 1117 IleValSerGlnCluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGly 1136  
Db 299 ATCGTGTCCAGAGCCCATCTGTTTGTACTGCAGCATTTGCTGAGAACAATGCCATATGA 358  
QY 1137 AspAsnSerArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIle 1156  
Db 359 GACAACAGCCGGTGTGTGCACAGGAAGAGATCGTGAGGCGCAGCAAGAGGCCCAACATA 418  
QY 1157 HisHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThr 1176  
Db 419 CATGCTTCATCAGTCACTGCTCCCTAATAATATAGCACTAAAGTAGGAGACAAAGAACT 478  
QY 1177 GlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnPro 1196  
Db 479 CAGCTCTCTGGTGGCCAGAAACAAACGCATGGCCATAGCTGCGCCCTGTTAGACAGCCT 538  
QY 1197 HisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysVal-Va 1216  
Db 539 CATATT--GCTTTGGATGAGCCAGCTCAGCTCTGATACAGAAAGTCAAAAGGTGCT 596  
QY 1216 lGlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgIle 1236  
Db 597 CCAAGAAGCCCTGGCAAAAGCCAGAGAAGGC---GACCTGCGATGTGTGTGTCCCGCTG 653  
QY 1236 uSerThrIleGlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHi 1256  
Db 654 -TCCACCATCCGAATGAGA--CTTATAGTGGTGTTCAGAAATGGCAGAGATCCAGGGGC 710
```

```
QY 1256 sGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSer 1271  
Db 711 ATGACCATCAGAGGTGCTGGACAGAAGS---CTCTATTTCATCA 753  
RESULT 10  
BG293345 602390738F1 NIH_MGC_94 726 bp mRNA linear EST 21-FEB-2001  
LOCUS BG293345 Mus musculus cDNA clone IMAGE:4502552 5',  
DEFINITION mRNA sequence.  
ACCESSION BG293345 GI:13052943  
VERSION BG293345.1  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 726)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: The Cepko Laboratory  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10371 row: 1 column: 09  
High quality sequence stop: 658.  
FEATURES  
source  
1..726  
Location/Qualifiers  
/organism="Mus musculus"  
/db_xref="taxon:10090"  
/clone="IMAGE:4502552"  
/clone_lib="NIH_MGC_94"  
/tissue_type="retina"  
/lab_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;  
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 3.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH_MGC Library."  
BASE COUNT 189 a 220 c 190 g 127 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 8.07e-75 Length: 726  
Score: 916.00 Matches: 185  
Percent Similarity: 94.66% Conservative: 10  
Best Local Similarity: 89.81% Mismatches: 9  
Query Match: 14.15% Indels: 2  
DB: 10 Gaps: 0  
US-09-672-725C-23 (1-1281) x BG293345 (1-726)  
QY 1068 LeuAlaLeuValGlySerSerglyCysGlyLysSerThrValValGlnLeuGluArg 1087  
Db 2 CTGGCCCTGGTGGCAGCAGTGGCTGGGGAGAGACACAGTGTCCAGTGTCTCGACGCG 61  
QY 1088 PheTyrAspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsn 1107  
Db 62 TTCATGACCCCATGGCTGGTGCATGCTCTTAGTAGTGTCAAGAAGCAAGAACTCAAT 121  
QY 1108 ValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCys 1127  
Db 122 GTCCAGTGGCTCCGACATCACTGGGCAATGTGCCAGGAACCACTCTCTTTGATGTC 181  
QY 1128 SerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgValValSerHisGluGluIle 1147  
Db 182 AGCATCGCAGAGACATCGCCTATGGAGACAACAGCCGGTCTGCTCATGATGAGATT 241
```

Qy 1148 MetGlnAlaLysGluAlaAsnIleHisPheIleGluThrLeuProGluLysTyr 1167
: : : : :
Db 242 GTGAGGGCAGCAAGGAGGCGCAACATCCACCCCTTCATCAGAGCGTCCGCCAAAATAT 301
Qy 1168 AsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGlnLysGlnArgIleAla 1187
: : : : :
Db 302 AACACAGATGAGGAGCAAGGAGGAGCGTCTCTGGGGCCAGAGGAGGATGGCC 361
Qy 1188 IleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAla 1207
: : : : :
Db 362 ATCGCCGAGCCCTCATCAGACCGCTCGGGTCTCTACTGCTGATGAGGACCGCTCAGCT 421
Qy 1208 LeuAspThrGluSerGluLysValValGlnGlnAlaLeuAspLysAlaArgGluGlyArg 1227
: : : : :
Db 422 CTGGATACTGAGAGTGAAGGTTGTCCAGGAAGCACTGGACAAAGCCAGGAGGCGCG 481
Qy 1228 ThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAla-AspLeuIleValVa 1247
: : : : :
Db 482 ACCTGCATTCGATCGCTCACCCTGTCCACCATCCAGACCGCGCACTTGATCGTGGT 541
Qy 1247 lPheGlnAsnGlyLysValLysGluHisGlyThr-HisGlnGlnLeuLeuAlaGlnLysG 1267
: : : : :
Db 542 GATTGAGACGGCAAGTCAAGGAGCAGCGCACCCACACGACGCTGCTGGCGAGGAGG 601
Qy 1267 lylleTyrPheSer 1271
: : : : :
Db 602 GCATCTATTTCTCA 615
RESULT 11
BM053000
LOCUS
DEFINITION
169a07.y3 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
cDNA 5' similar to SW-MDR1_HUMAN P08183 MULTIDRUG RESISTANCE
PROTEIN 1 ;, mRNA sequence.
ACCESSION
BM053000
VERSION
BM053000.1 GI:1680894
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 587)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Secorce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
J., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
High quality sequence stop: 437.
FEATURES
source
Location/Qualifiers
1..587
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;

Site_2: Sal 1; Starting library constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an Ecot of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 142 a 138 c 161 g 145 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 2,21e-74 Length: 587
Score: 910.00 Matches: 175
Percent Similarity: 96.34% Conservative: 9
Best Local Similarity: 91.62% Mismatches: 7
Query Match: 14.06% Indels: 0
DB: 10 Gaps: 0

US-09-672-725C-23 (1-1281) x BM053000 (1-587)

Qy 950 MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu 969
: : : : :
Db 3 ATGATGATATTTTCCCTATCTGGATGTTTCCGGTTTGGAGCCTACTTGTGGCATAA 62

Qy 970 PheMetAsnPheGlnValLeuLeuValPheSerAlaIleValPheGlyAlaMetAla 989
: : : : :
Db 63 CTCATGAGCTTCAGATGTTCTGTAGTATTTTCAGCTGTTGTCTTGGTCCATGCC 122

Qy 990 ValGlyGlnValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis 1009
: : : : :
Db 123 GTGGGCAAGTCAGTTTCATTGCTCTGACTATGCCAAGCCAAAATATCAGCAGCCAC 182

Qy 1010 ValIleMetIleIleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys 1029
: : : : :
Db 183 ATCATCATCATCTGAAAAAACCCCTTTGATTGACAGCTACAGCAGCGAGGCGCTAATG 242

Qy 1030 ProAsnThrLeuGluGlyAsnValThrPheAsnGlnValValPheAsnTyrProThrArg 1049
: : : : :
Db 243 CCGACACATTTGGAAGAAATGTCACATTTGGTGAAGTTGTATTCACTATCCACCCGA 302

Qy 1050 ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysLysGlyIleThrLeuAla 1069
: : : : :
Db 303 CCGGACATCCCATGCTTCAGGAGCTGAGCCTGGAGGTGAAGAAGGCCAGCGTGGCT 362

Qy 1070 LeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyr 1089
: : : : :
Db 363 CTGGTGGGAGCAGTGGCTGTGGGAAGACAGTGGTCCAGCTCTCTGAGGCGGTCTAC 422

Qy 1090 AspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGln 1109
: : : : :
Db 423 GACCCCTTGGCAGGGAAGTGTCTGTTGATGCCAAGAAATAAACCGACTGAATGTTTAC 482

Qy 1110 TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle 1129
: : : : :
Db 483 TGGCTCCGAGCACACCTGGGCATCGTGTCCAGGAGGCCATCCTCTTTTGTGACTGCAGCAT 542

Qy 1130 AlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
: : : : :
Db 543 GCTGAGAACATTTGCCTATGGAGACACACGCCGN 575

RESULT 12

BM080311

LOCUS

DEFINITION

BM080311

ACCESSION

BM080311.1

628 bp mRNA linear EST 26-JAN-2001
H3052B06-5 NTA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3052B06 5', mRNA sequence.
GI:12562879

[illegible]

Research 6.791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library." *Journal of Molecular Biology*, Vol. 175, No. 1, 1984, pp. 1-10.

Qy 950 MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu 969
|||||
|||||

Alignment Scores:
Pred. No.: 1.87e-72 Length: 563

Score: 889.00 Matches: 171
Percent Similarity: 96.26% Conservative: 9
Best Local Similarity: 91.44% Mismatches: 7
Query Match: 13.73% Indels: 0
DB: 10 Gaps: 0

US-09-672-725C-23 (1-1281) x B1964218 (1-563)

Qy 950 MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu 969
Db 3 ATGATGATATTTTCCATGCTGGATGTTCCGGTTTGGAGCCTACTTGGTGGCACAATAA 62

Qy 970 PheMetAsnPheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAla 989
Db 63 CTCATGAGCTTTGAGGATGTTCTGTTAGTATTTTCAGCTGTTGCTTGGTGGCATTGCC 122

Qy 990 ValGlyGlnValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis 1009
Db 123 GTGGGCAAGTCAGTTCATTTGCTGCTGACTATGCCAAGCCAAATATCAGACGCCAC 182

Qy 1010 ValIleMetIleGlyLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys 1029
Db 183 ATCATCATCATCTTGAATAAACCCCTTTGATTGACAGCTACAGCAGCGCCCTAATG 242

Qy 1030 ProAsnThrLeuGluClyAsnValThrPheAsnGluValValPheAsnTyrProThrArg 1049
Db 243 CCGAACACATTTGAGGAATATCAGATTTGGTGAAGTTGTATTCAACTATCCACCGCA 302

Qy 1050 ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyClnThrLeuAla 1069
Db 303 CCGGACATCCCATGCTTCCAGGACTGAGCCTGGAGGTGAAGAAGGGCCAGCGCTGGCT 362

Qy 1070 LeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuGluArgPheTyr 1089
Db 363 CTGTTGGGACAGTGGCTGTGGGAGACACAGTGGTCCAGCTCCTGGAGCGGTCTTAC 422

Qy 1090 AspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGln 1109
Db 423 GACCCCTTGGCAGGAAAGTCTGCTTGTATGGCAAGAAATAAAGCGACTGAATGTTCCAG 482

Qy 1110 TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle 1129
Db 483 TGCTCCGAGACACATCGGGCATCGTGTCCAGGAGCCATCCTGTTTGACTGTCAGCATT 542

Qy 1130 AlaGluAsnIleAlaTyrGly 1136
Db 543 GCTGAGAACATTTGCCTATGGA 563

RESULT 15
BM052878
LOCUS
DEFINITION
ie67e03.y3 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
CDNA 5', similar to SK-MDL1_HUMAN P08183 MULTIDRUG RESISTANCE
PROTEIN 1 ;, mRNA sequence.

ACCESSION
BM052878
VERSION
BM052878.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Narra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute

Dept. of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
High quality sequence stop: 430.

FEATURES
source
1..563
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an Ecot of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 136 a 132 c 153 g 142 t
ORIGIN

Alignment Scores:
Pred. No.: 1,87e-72 Length: 563
Score: 889.00 Matches: 171
Percent Similarity: 96.26% Conservative: 9
Best Local Similarity: 91.44% Mismatches: 7
Query Match: 13.73% Indels: 0
DB: 10 Gaps: 0

US-09-672-725C-23 (1-1281) x BM052878 (1-563)

Qy 950 MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu 969
Db 3 ATGATGATATTTTCCATGCTGGATGTTCCGGTTTGGAGCCTACTTGGTGGCACAATAA 62

Qy 970 PheMetAsnPheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAla 989
Db 63 CTCATGAGCTTTGAGGATGTTCTGTTAGTATTTTCAGCTGTTGCTTGGTGGCATTGCC 122

Qy 990 ValGlyGlnValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis 1009
Db 123 GTGGGCAAGTCAGTTCATTTGCTGCTGACTATGCCAAGCCAAATATCAGACGCCAC 182

Qy 1010 ValIleMetIleGlyLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys 1029
Db 183 ATCATCATCATCTTGAATAAACCCCTTTGATTGACAGCTACAGCAGCGCCCTAATG 242

Qy 1030 ProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArg 1049
Db 243 CCGAACACATTTGAGGAATATCAGATTTGGTGAAGTTGTATTCAACTATCCACCGCA 302

Qy 1050 ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyClnThrLeuAla 1069
Db 303 CCGGACATCCCATGCTTCCAGGACTGAGCCTGGAGGTGAAGAAGGGCCAGCGCTGGCT 362

Qy 1070 LeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuGluArgPheTyr 1089
Db 363 CTGTTGGGACAGTGGCTGTGGGAGACACAGTGGTCCAGCTCCTGGAGCGGTCTTAC 422

```
Oy 1090 AspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGln 1109
Db 423 GACCCCTGGCAGGGAAGTCTGCTTGAATGGAATAAAGCAGCTGAATGTTTCAG 482
Oy 1110 TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle 1129
Db 483 TGGCTCCGAGCACACCTGGGCATCGTGCCAGGAGCCCATCCTGTTTGACTGCAGCATT 542
Oy 1130 AlaGluAsnIleAlaTyrGly 1136
Db 543 GCTGAGAACAATTGCCTATGGA 563
```

Search completed: November 7, 2002, 07:28:59
Job time : 2251.96 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 17:45:54 ; Search time 31.4943 Seconds
(without alignments)
4517.824 Million cell updates/sec

Title: US-09-672-725C-25
Perfect score: 6473
Sequence: 1 MDPEGGRKGAERNFWMKGK.....LLAQKGIYFMSVQAGAKR 1281

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_032802.*
- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
 - 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
 - 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
 - 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
 - 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
 - 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
 - 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
 - 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
 - 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
 - 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
 - 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
 - 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
 - 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
 - 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
 - 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
 - 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
 - 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
 - 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
 - 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
 - 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
 - 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
 - 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	6473	100.0	1281	AAE00309	Dog P-glycoprotein
2	6468	99.9	1281	AAE00308	Dog P-glycoprotein
3	6466	99.9	1281	AAE00310	Dog P-glycoprotein
4	6463	99.8	1281	AAE00303	Dog (PGP) P-glycop
5	6425.5	99.3	1280	AAE00304	Dog P-glycoprotein
6	6425.5	99.3	1280	AAE00304	Dog P-glycoprotein
7	5866	90.6	1280	AAE00304	Human multidrug re
8	5866	90.6	1280	AAE00304	Human multidrug re
9	5866	90.6	1280	AAE00304	Human multidrug re
10	5862	90.6	1280	AAE00304	Human multidrug re
11	5859.5	90.5	1283	AAE00304	Cynomolgous monke

12	5857	90.5	1280	21	AAE00309	Human G185V mutant
13	5857	90.5	1280	22	AAE00309	Human P-glycoprote
14	5857	90.5	1280	22	AAE00309	Human P-glycoprote
15	5854	90.4	1280	8	AAE00309	Sequence encoded b
16	5851	90.4	1280	11	AAE00309	Protein encoded by
17	5849.5	90.4	1279	22	AAE00309	Human P-glycoprote
18	5849.5	90.4	1279	22	AAE00309	Human P-glycoprote
19	5844	90.3	1280	15	AAE00309	Human P-glycoprote
20	5841	90.2	1280	14	AAE00309	Sequence encoded b
21	5841	90.2	1280	22	AAE00309	Human multidrug re
22	5840.5	90.2	1279	19	AAE00309	Mutated human P-gl
23	5830	90.1	1280	19	AAE00309	Wild-type human P
24	5824	90.0	1280	19	AAE00309	N-terminal single
25	5824	90.0	1280	19	AAE00309	C-terminal single
26	5818	89.9	1280	19	AAE00309	Human P glycoprote
27	5818	89.9	1280	19	AAE00309	Murine multidrug r
28	5621	86.8	1272	21	AAE00309	Rat multidrug resi
29	5611	86.7	1272	21	AAE00309	Rat multidrug resi
30	5301	81.9	1276	14	AAE00309	Mouse multidrug re
31	5301	81.9	1276	21	AAE00309	Murine multidrug r
32	5291.5	81.7	1275	21	AAE00309	Rat mdrlb2 (multis
33	5291.5	81.7	1275	22	AAE00309	Rat mdrlb2 multidr
34	2809.5	43.4	1275	20	AAE00309	H. contortus PGP-A
35	2694	41.6	1313	22	AAE00309	Drosophila melanog
36	2555	39.5	1279	22	AAE00309	Drosophila melanog
37	2513	38.8	1349	20	AAE00309	Multiple drug resi
38	2491	38.5	1302	22	AAE00309	Drosophila melanog
39	2481	38.3	1334	20	AAE00309	Protein encoded by
40	2446	37.8	1408	19	AAE00309	Multiple drug resi
41	2424	37.4	1286	21	AAE00309	Arabidopsis thalia
42	2401.5	37.1	1333	22	AAE00309	P. chrysogenum ABC
43	2397	37.0	1243	21	AAE00309	Arabidopsis thalia
44	2388	36.9	1320	22	AAE00309	Drosophila melanog
45	2382	36.8	1254	21	AAE00309	A. thaliana ATPAC,

ALIGNMENTS

RESULT 1	
AAE00309	AAE00309 standard; Protein; 1281 AA.
ID	AAE00309 standard; Protein; 1281 AA.
XX	AC
XX	AAE00309;
XX	13-JUN-2001 (first entry)
DT	Dog P-glycoprotein (PGP) allelic variant (Genotype B).
DE	Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
KW	MDRL; drug bioavailability; transgenic animal; genetic model.
KW	MDRL; drug bioavailability; transgenic animal; genetic model.
XX	Canis familiaris.
OS	Canis familiaris.
XX	Key
FH	Location/Qualifiers
FT	Misc-difference 25
FT	Misc-difference 197
FT	Misc-difference 197
FT	Misc-difference 197
FT	Misc-difference 197
XX	WO200123540-A2.
XX	05-APR-2001.
XX	28-SEP-2000; 2000WO-US26767.
XX	28-SEP-1999; 99US-0156510.
XX	(GENT-) GENTEST CORP.
XX	Stockert PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX	WPI; 2001-235373/24.

DR N-PSDB; AAD03505.
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
XX
PS Claim 17; Page 99-102; 11pp; English.
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDR1) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
CC This sequence is also referred as Genotype B protein. The PGP
CC enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.
XX
SQ Sequence 1281 AA;

Query Match 100.0%; Score 6473; DB 22; Length 1281;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEGGRGSAKFNWKKSKKSKKKEKKPTVSTFAMFYSNWLDRLYMLVGTMAII 60
DB 1 MDEGGRGSAKFNWKKSKKSKKKEKKPTVSTFAMFYSNWLDRLYMLVGTMAII 60

QY 61 HGAALPLMLLVFGNMTDSFANAGISRNKTFPVIINESITNTNQHPINLHEEMTYAYYY 120
DB 61 HGAALPLMLLVFGNMTDSFANAGISRNKTFPVIINESITNTNQHPINLHEEMTYAYYY 120

QY 121 SGIGAGVLVAAYIQVSWFCLAAAGROILKIRKOFFHAIMRQEGFVDHVGELNTRLTDD 180
DB 121 SGIGAGVLVAAYIQVSWFCLAAAGROILKIRKOFFHAIMRQEGFVDHVGELNTRLTDD 180

QY 181 VSKINEGIGDKIGMFFQSTATFTTCFIVGFTGKWLTVILAIASPVGLSAAIWAAILSS 240
DB 181 VSKINEGIGDKIGMFFQSTATFTTCFIVGFTGKWLTVILAIASPVGLSAAIWAAILSS 240

QY 241 FTDKELLAYAKAGAAEEVLAIAIRTVIAFGGQKKELERYKNLEAKGIGIKKAITANIS 300
DB 241 FTDKELLAYAKAGAAEEVLAIAIRTVIAFGGQKKELERYKNLEAKGIGIKKAITANIS 300

QY 301 IGAAFLIIYASVALAFWGTSTVLVLSSEYSIGQVLVFFSVLIGAFSIGQASPSIEAFANA 360
DB 301 IGAAFLIIYASVALAFWGTSTVLVLSSEYSIGQVLVFFSVLIGAFSIGQASPSIEAFANA 360

QY 361 RGAAYEIFKIIDNKPSIDSYSGSHKPNIKGNLEFKNWHPFSYPSRKEVKILGLNLKVQ 420
DB 361 RGAAYEIFKIIDNKPSIDSYSGSHKPNIKGNLEFKNWHPFSYPSRKEVKILGLNLKVQ 420

QY 421 SGQTVLVGNSCGKSTTVQLMORLYDPTDGMVCIDGQDIRTINVRHLREITGVVSQBPV 480
DB 421 SGQTVLVGNSCGKSTTVQLMORLYDPTDGMVCIDGQDIRTINVRHLREITGVVSQBPV 480

QY 481 LPATTIAENIRYGRNVMTDETEKAVKANAYDFIMKLPNKPDTLVGERGQALSGGQKQR 540
DB 481 LPATTIAENIRYGRNVMTDETEKAVKANAYDFIMKLPNKPDTLVGERGQALSGGQKQR 540

QY 541 IAIARALVRNPKILLDDEATSDALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
DB 541 IAIARALVRNPKILLDDEATSDALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVI 600

QY 601 AGDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELEENATGESKESDALEMPKDS 660
DB 601 AGDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELEENATGESKESDALEMPKDS 660

QY 661 GSSLIKRRSTRSRSHAPQGDRLTKEDLNENVPVSWFRILKLNSTWEPYFVVGIFCA 720
DB 661 GSSLIKRRSTRSRSHAPQGDRLTKEDLNENVPVSWFRILKLNSTWEPYFVVGIFCA 720

QY 721 IINGGLQPAFSIIFSRIGTITRDEDPETKRONSNFSLVFLVLGIISFITFFLOQFTFG 780
DB 721 IINGGLQPAFSIIFSRIGTITRDEDPETKRONSNFSLVFLVLGIISFITFFLOQFTFG 780

QY 781 KAGEILTFRLYRMVFRSMLRQDYSWFDKNTTTCALTTRLANDAAQVKAIGSRLAVITQ 840
DB 781 KAGEILTFRLYRMVFRSMLRQDYSWFDKNTTTCALTTRLANDAAQVKAIGSRLAVITQ 840

QY 841 NIANLGTGIIISLYGWQLTLLLAIVPIITAIAGVVMKMLSGQALKDKKELEGAGKIAT 900
DB 841 NIANLGTGIIISLYGWQLTLLLAIVPIITAIAGVVMKMLSGQALKDKKELEGAGKIAT 900

QY 901 EATENFTVYSLTREOKFEYMAQSLQVYPYNSLRKAHIFGVFSITQAMMYFSYAGCFR 960
DB 901 EATENFTVYSLTREOKFEYMAQSLQVYPYNSLRKAHIFGVFSITQAMMYFSYAGCFR 960

QY 961 FGAYLVANEFMNFQDVLVLSAIVFGAMAVQVSSPADYAKAKVSAAHVIMIIKSPLI 1020
DB 961 FGAYLVANEFMNFQDVLVLSAIVFGAMAVQVSSPADYAKAKVSAAHVIMIIKSPLI 1020

QY 1021 DSYSPHGLKNTLEGNVTFNEVFNYPTRDIPVLOGLSLEVKKGQTLALVSSGCGKST 1080
DB 1021 DSYSPHGLKNTLEGNVTFNEVFNYPTRDIPVLOGLSLEVKKGQTLALVSSGCGKST 1080

QY 1081 WVOLLERFYDPLAGSLVLDGKEIKHLNVQWLAHGLTVSQEPILFDCSIAENIAYGDSNR 1140
DB 1081 WVOLLERFYDPLAGSLVLDGKEIKHLNVQWLAHGLTVSQEPILFDCSIAENIAYGDSNR 1140

QY 1141 VVSHEETMQAAKEANITHFTIPEKYNTVRGDKGTQSGGQKORIAIARALVRQPHILL 1200
DB 1141 VVSHEETMQAAKEANITHFTIPEKYNTVRGDKGTQSGGQKORIAIARALVRQPHILL 1200

QY 1201 LDEATSDLTSEKVVQOEALDKAREGTCIVIAHRLSTIQNADLIYVFONGKVEHGTQ 1260
DB 1201 LDEATSDLTSEKVVQOEALDKAREGTCIVIAHRLSTIQNADLIYVFONGKVEHGTQ 1260

QY 1261 QLLAQKGIYFSMVSQVQAGAKR 1281
DB 1261 QLLAQKGIYFSMVSQVQAGAKR 1281

RESULT 2
AAE00308
ID AAE00308 standard; Protein; 1281 AA.
XX
AC AAE00308;
XX
DT 13-JUN-2001 (first entry)
XX
DE Dog P-glycoprotein (PGP) allelic variant (Genotype A).
XX
KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
KW MDR1; drug bioavailability; transgenic animal; genetic model.
XX
OS Canis familiaris.
XX
XX Key Location/Qualifiers
FT Misc-difference 197 /note= "His of GenotypeC substituted by Gln"
XX
XX WO200123540-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26767.

XX 28-SEP-1999; 99US-0156510.
PR (GENF-) GENTEST CORP.
PA
XX
XX Stocker FU, Stelmel-crespi DT, Crespi CL, Reif TC, Patten CJ;
PI WPI; 2001-235373/24.
DR N-PSDB; RAD03504.
XX
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
XX Claim 17; Page 91-93; 111pp; English.
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDR1) and nucleic acids encoding thereof.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
CC This sequence is also referred as Genotype A protein. The PGP
CC enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.
XX
XX Sequence 1281 AA;
50

Query Match 99.9%; Score 6468; DB 22; Length 1281;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1280; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MDEGGRKGSAAKFNKMGKSKKKKKKPTVSTFAMFRYSNWLDRMLVGTMAAII 60
Db 1 MDEGGRKGSAAKFNKMGKSKKKKKKPTVSTFAMFRYSNWLDRMLVGTMAAII 60
Qy 61 HGAALPLMLVFGNMTDSFANAGISRNKTPPVIINESITNTQHFINHLEEMTTAYYY 120
Db 61 HGAALPLMLVFGNMTDSFANAGISRNKTPPVIINESITNTQHFINHLEEMTTAYYY 120
Qy 121 SGIGAGVLVAAYIOVSEFWCLAAGROILKIRKQPFHAIMROEIGWFDVHDVGNLRLTDD 180
Db 121 SGIGAGVLVAAYIOVSEFWCLAAGROILKIRKQPFHAIMROEIGWFDVHDVGNLRLTDD 180
Qy 181 VSKINEGIDKIGMFFQSIATFTFTGFTVGRGKMLFLVILAIISPVLGSAAIWAKILSS 240
Db 181 VSKINEGIDKIGMFFQSIATFTFTGFTVGRGKMLFLVILAIISPVLGSAAIWAKILSS 240
Qy 241 FTDKELLAYAKAGAAVEVLAAIRTVIAFGQKKELERYKNLEEAIGIKKAITANIS 300
Db 241 FTDKELLAYAKAGAAVEVLAAIRTVIAFGQKKELERYKNLEEAIGIKKAITANIS 300
Qy 301 IGAAPFLIYASALAFWGTSLVLSSEYSIGQVLTVFFSVLIGAFSIGQASPSIEAFANA 360
Db 301 IGAAPFLIYASALAFWGTSLVLSSEYSIGQVLTVFFSVLIGAFSIGQASPSIEAFANA 360
Qy 361 RGAAYEIFKIIDNKPISDSYKSGHKPDNKGKLEFNHVSFYSRKEVKILGLNLKVQ 420
Db 361 RGAAYEIFKIIDNKPISDSYKSGHKPDNKGKLEFNHVSFYSRKEVKILGLNLKVQ 420
Qy 421 SGQTVLVGNSGGCKSTTVQLMORLYDPTGMVCIDQDRTINVRHLREITGVVSOEPV 480
Db 421 SGQTVLVGNSGGCKSTTVQLMORLYDPTGMVCIDQDRTINVRHLREITGVVSOEPV 480

QY 481 LFATTIAENIRYGRNVMTDEIEKAYKEANAYDFIMKLPNKFTLIVGERCAQLSGGQKOR 540
Db 481 LFATTIAENIRYGRNVMTDEIEKAYKEANAYDFIMKLPNKFTLIVGERCAQLSGGQKOR 540
QY 541 IATARALVRNPKILLDLDEATSDALDTESEAVVQVALDKARKGRITTVIAHRLSVRNADVI 600
Db 541 IATARALVRNPKILLDLDEATSDALDTESEAVVQVALDKARKGRITTVIAHRLSVRNADVI 600
QY 601 AGFDDGVIVEKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGESKESDALEMPKDS 660
Db 601 AGFDDGVIVEKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGESKESDALEMPKDS 660
QY 661 GSSLIKRRSRHSIHAPOGDRKLTGKEDLNENVPVSEWRILKLNSTWPEVVGIFCA 720
Db 661 GSSLIKRRSRHSIHAPOGDRKLTGKEDLNENVPVSEWRILKLNSTWPEVVGIFCA 720
QY 721 IINGGLQPAFSIIISRIIGIFTRDEPETKRONSNMFSVLFLVLGIISITITFFLOGFTFG 780
Db 721 IINGGLQPAFSIIISRIIGIFTRDEPETKRONSNMFSVLFLVLGIISITITFFLOGFTFG 780
QY 781 KAGEILTKRLRYMFRSMLRQDVSWFDDPKNTTGALTTLANDAAQVKAIGSLRAVITQ 840
Db 781 KAGEILTKRLRYMFRSMLRQDVSWFDDPKNTTGALTTLANDAAQVKAIGSLRAVITQ 840
QY 841 NIANLGTGIIISLIYGWLTLLLAIVPIIATAGVVMKMLSGOALKDKKELEGAGKIAT 900
Db 841 NIANLGTGIIISLIYGWLTLLLAIVPIIATAGVVMKMLSGOALKDKKELEGAGKIAT 900
QY 901 EAIENERTVYSLTREQKFEYMAQSLQVPRNSLRKAHIFGVVSFSTQAMMYFSYAGCFR 960
Db 901 EAIENERTVYSLTREQKFEYMAQSLQVPRNSLRKAHIFGVVSFSTQAMMYFSYAGCFR 960
QY 961 FGAYLVANEFMNFQDVLVFSATVFGAMAVGVSSPAPDYAKAKVSAAHVIMIEKSPLI 1020
Db 961 FGAYLVANEFMNFQDVLVFSATVFGAMAVGVSSPAPDYAKAKVSAAHVIMIEKSPLI 1020
QY 1021 DSYSPHGLKPNLTLEGNVTNEVFNYPTRPDIPVLOGLSLEVKKGOTLALVGSSGCGKST 1080
Db 1021 DSYSPHGLKPNLTLEGNVTNEVFNYPTRPDIPVLOGLSLEVKKGOTLALVGSSGCGKST 1080
QY 1081 VQQLERFYDPLAGSVLIDGKEIKHLNVQWLAHILGIVSQEPIFLDCSTAEINAYGDNRSR 1140
Db 1081 VQQLERFYDPLAGSVLIDGKEIKHLNVQWLAHILGIVSQEPIFLDCSTAEINAYGDNRSR 1140
QY 1141 VVSHEEIMQAANKANTHFIETLPEKYNTRVGDKGTLQSGGQKQRIATARALVRPHILL 1200
Db 1141 VVSHEEIMQAANKANTHFIETLPEKYNTRVGDKGTLQSGGQKQRIATARALVRPHILL 1200
QY 1201 LDEATSDALDTESEKVVQEAALDKAREGRTCTIVIAHRLSTIQNADLIIVFONGVKKEGTHQ 1260
Db 1201 LDEATSDALDTESEKVVQEAALDKAREGRTCTIVIAHRLSTIQNADLIIVFONGVKKEGTHQ 1260
QY 1261 QLLAQKGIYFSVMVQAGAKR 1281
Db 1261 QLLAQKGIYFSVMVQAGAKR 1281
RESULT 3
AAE00310
ID AAE00310 standard; Protein; 1281 AA.
XX
AC AAE00310;
XX
XX 13-JUN-2001 (first entry)
DT
XX
DE Dog P-glycoprotein (PGP) allelic variant (Genotype D).
XX
KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
MDRI; drug bioavailability; transgenic animal; genetic model.
XX
OS Canis familiaris.
XX
FH Key Location/Qualifiers

Misc-difference 25 /note= "Asn of GenotypeC substituted by Lys"
Misc-difference 197 /note= "His of GenotypeC substituted by Gln"
Misc-difference 329 /note= "Ser of GenotypeC substituted by Thr"
Misc-difference 1148 /note= "Met of GenotypeC substituted by Val"
WO200123540-A2.
05-APR-2001.
28-SEP-2000; 2000WO-US26767.
28-SEP-1999; 99US-0156510.
(GENT-) GENTEST CORP.
Stocker PJ, Stelmel-crespi DT, Crespi CL, Reif TC, Patten CJ;
WPI; 2001-235373/24.
N-PSDB; AAD03506.
New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
for determining the bioavailability of drugs and for screening for dog
PGP inhibitors -
Claim 17; Page 108-110; 111pp; English.
The invention relates to dog P-glycoprotein (PGP) also referred
as multidrug transporter (MDR1) and nucleic acids encoding them.
The invention also includes fragments and biologically functional
variants of dog P-glycoprotein. PGP and their nucleic acids are
useful for determining the bioavailability of drugs and for
screening PGP inhibitors. They are useful for the diagnosis and
treatment of conditions characterised by PGP activity, by
reducing or increasing PGP activity in a cell. PGP nucleic acids
are used as oligonucleotide probes. Complements of PGP nucleic
acids are useful as antisense oligonucleotides, to induce a PGP
'knockout' phenotype. They are used to prepare a non-human
transgenic animal, which are valuable as genetic models for
human diseases.
CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
CC This sequence is also referred as Genotype D protein. The PGP
CC enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.
XX
SQ Sequence 1281 AA;
Query Match 99.9%; Score 6466; DB 22; Length 1281;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1279; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MDPEGGRKSAENFWKMGKSKKKEKKKPTVSTFAMFRYSNWLDRMLVGMTMAAII 60
Dy 1 MDPEGGRKSAENFWKMGKSKKKEKKKPTVSTFAMFRYSNWLDRMLVGMTMAAII 60
Qy 61 HGAALPLMLVFGNMTDSFANAGISRNKTPPVLINESITNNTQHFHNHLEEMTTVAYYY 120
Dy 61 HGAALPLMLVFGNMTDSFANAGISRNKTPPVLINESITNNTQHFHNHLEEMTTVAYYY 120
Qy 121 SGIGAGVLVAAYIQVSWFLAACRQILKIRKQFFHAIMRQEIQWDFVDVHVGELNRLTDD 180
Dy 121 SGIGAGVLVAAYIQVSWFLAACRQILKIRKQFFHAIMRQEIQWDFVDVHVGELNRLTDD 180
Qy 181 VSKINEGDKIGMFFOSIAFTFTGFTVGTGRGWKLTVLILASPVLGSAATWAKILSS 240
Dy 181 VSKINEGDKIGMFFOSIAFTFTGFTVGTGRGWKLTVLILASPVLGSAATWAKILSS 240
Qy 241 FTDKELLAYAKAGAVAEVLAARTVIAFGOKKELERYNKNLEEAAGIGIKKAITANIS 300
Dy 241 FTDKELLAYAKAGAVAEVLAARTVIAFGOKKELERYNKNLEEAAGIGIKKAITANIS 300

Db 241 FTDKELLAYAKAGAVAEVLAARTVIAFGOKKELERYNKNLEEAAGIGIKKAITANIS 300
Qy 301 IGA AFLLIYASYALAFWYGTSLVLSSEYSISQVLTVFFSVLIGAFSGQASPSIEAFANA 360
Db 301 IGA AFLLIYASYALAFWYGTSLVLSSEYSISQVLTVFFSVLIGAFSGQASPSIEAFANA 360
Qy 361 RGAAYEIFKIIDNKPISDSYKSGHKPDNKGNEFNHVSFSPSRKEVKILGKLNKQV 420
Db 361 RGAAYEIFKIIDNKPISDSYKSGHKPDNKGNEFNHVSFSPSRKEVKILGKLNKQV 420
Qy 421 SGQTVLVGNSGCGKSTTVQLMORLYDPTDGMWCIDQDRTINVRHLREITGVVSEPV 480
Db 421 SGQTVLVGNSGCGKSTTVQLMORLYDPTDGMWCIDQDRTINVRHLREITGVVSEPV 480
Qy 481 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQSGGQKOR 540
Db 481 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQSGGQKOR 540
Qy 541 IAIARALVRNPKILLDEATSAIDTSEAVVQVALDKARKGRRTIVIAHRLSTVRNADVI 600
Db 541 IAIARALVRNPKILLDEATSAIDTSEAVVQVALDKARKGRRTIVIAHRLSTVRNADVI 600
Qy 601 AGFDDGVIVEKGNHDELMKEGIYFKLVMTOTRGNEIELENATGESKESDALEMSPKDS 660
Db 601 AGFDDGVIVEKGNHDELMKEGIYFKLVMTOTRGNEIELENATGESKESDALEMSPKDS 660
Qy 661 GSSLIKRSTRRSIHAPOQDRKLGTRKEDLNENPPVSWRILKLNSTWEPYFVVGIFCA 720
Db 661 GSSLIKRSTRRSIHAPOQDRKLGTRKEDLNENPPVSWRILKLNSTWEPYFVVGIFCA 720
Qy 721 IINGLOPAPSIIFRSIIGITRDEDPETKQNSNMSEVFLVLGIIISFTFFLQGTFFG 780
Db 721 IINGLOPAPSIIFRSIIGITRDEDPETKQNSNMSEVFLVLGIIISFTFFLQGTFFG 780
Qy 781 KAGELLTKRLYMWFRSMRDVSWFDDPKNTTGALTTRLANDAAQVKAIGSRNAVITQ 840
Db 781 KAGELLTKRLYMWFRSMRDVSWFDDPKNTTGALTTRLANDAAQVKAIGSRNAVITQ 840
Qy 841 NIANLGTGIIISLIYGWOLTLALLAIVPIIAIAGVVENKMLSGQALDKKKELEGAGKIAT 900
Db 841 NIANLGTGIIISLIYGWOLTLALLAIVPIIAIAGVVENKMLSGQALDKKKELEGAGKIAT 900
Qy 901 EAIENFRVVSILTREQKFEYMAQSLQVPRNSLRKAHIFGVFSFISITQAMMYFYAGCFR 960
Db 901 EAIENFRVVSILTREQKFEYMAQSLQVPRNSLRKAHIFGVFSFISITQAMMYFYAGCFR 960
Qy 961 FGAYLVANEFMNFQDVLVFSIAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMIEKSPLI 1020
Db 961 FGAYLVANEFMNFQDVLVFSIAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMIEKSPLI 1020
Qy 1021 DSYSPHGLKPNLTLEGNVTFNEVFNYPTRPDI PVLOGLSLEVKKGQTLALVGSSECGKST 1080
Db 1021 DSYSPHGLKPNLTLEGNVTFNEVFNYPTRPDI PVLOGLSLEVKKGQTLALVGSSECGKST 1080
Qy 1081 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWLAHGLGIVSQEPILFDCSIAENIAYGNSR 1140
Db 1081 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWLAHGLGIVSQEPILFDCSIAENIAYGNSR 1140
Qy 1141 VVSHEEIQAAKEANIHHFIETLPEKYNTRYGDKGTQLSGGQKQRIATARALVRPHILL 1200
Db 1141 VVSHEEIQAAKEANIHHFIETLPEKYNTRYGDKGTQLSGGQKQRIATARALVRPHILL 1200
Qy 1201 LDEATSAIDTSEKVVQBALDKAREGRTCTIVIAHRLSTIQNALDLYVFPQNKVEHGTQ 1260
Db 1201 LDEATSAIDTSEKVVQBALDKAREGRTCTIVIAHRLSTIQNALDLYVFPQNKVEHGTQ 1260
Qy 1261 QLLAQKGIYFSMVSVQAGAKR 1281
Db 1261 QLLAQKGIYFSMVSVQAGAKR 1281
RESULT 4
AAE00303

ID AAE00303 standard; Protein; 1281 AA.
 AC AAE00303;
 DT 13-JUN-2001 (first entry)
 XX Dog (Pgp) P-glycoprotein (genotype C) #1.
 DE Dog; P-glycoprotein; Pgp; multidrug transporter; MDRL;
 KW drug bioavailability; transgenic animal; genetic model.
 XX Canis familiaris.
 OS WO200123540-A2.
 PN 05-APR-2001.
 PD 28-SEP-2000; 2000WO-US26767.
 PF 28-SEP-1999; 99US-0156510.
 PR (GENT-) GENTEST CORP.
 PA Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 XX WPI; 2001-235373/24.
 DR N-PSDB; AAD03488.
 XX New dog P-glycoproteins (Pgp) and their encoding nucleic acids, useful
 PT for determining the bioavailability of drugs and for screening for dog
 PT Pgp inhibitors -
 PS Claim 17; Page 64-66; 11pp; English.
 CC The invention relates to dog P-glycoprotein (Pgp) also referred
 CC as multidrug transporter (MDRL) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. Pgp and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening Pgp inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by Pgp activity, by
 CC reducing or increasing Pgp activity in a cell. Pgp nucleic acids
 CC are used as oligonucleotide probes. Complements of Pgp nucleic
 CC acids are useful as antisense oligonucleotides, to induce a Pgp
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (Pgp) also referred
 CC as genotype C protein. The Pgp enzyme functions as an efflux
 CC pump exporting small molecules across the cell membrane. This
 CC enzyme is a member of the ABC transporter family.
 XX Sequence 1281 AA;
 SQ
 Query Match 99.8%; Score 6463; DB 22; Length 1281;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1279; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MDPEGGRGSAEKFNFKMGKSKKKEKKKPTVSTFAMPFYSNWLDRMLVGTMAAII 60
 DB 1 MDPEGGRGSAEKFNFKMGKSKKKEKKKPTVSTFAMPFYSNWLDRMLVGTMAAII 60
 QY 61 HGAALPLMLVFGNMTDSFANAGISRNTKTPVIINESITNTTOHFINHLEEMTTVAYY 120
 DB 61 HGAALPLMLVFGNMTDSFANAGISRNTKTPVIINESITNTTOHFINHLEEMTTVAYY 120
 QY 121 SGIGAGVLVAAYIQVSWFCLAAAGQILKIRKQFFHAIMRQIEIGWFDVHDVGEINLRLTDD 180
 DB 121 SGIGAGVLVAAYIQVSWFCLAAAGQILKIRKQFFHAIMRQIEIGWFDVHDVGEINLRLTDD 180
 QY 181 VSKINEGIGDKIGMFFOSIATFTFTGTVGRGWKLVILATSPVLGLSAAIWKILSS 240
 DB 181 VSKINEGIGDKIGMFFHSIATFTFTGTVGRGWKLVILATSPVLGLSAAIWKILSS 240

QY 241 FTDKELLAYAKAGAAVEVLAAIRTVIAFGGQKKELERYNKNLEEAAGIGIKKAITANIS 300
 DB 241 FTDKELLAYAKAGAAVEVLAAIRTVIAFGGQKKELERYNKNLEEAAGIGIKKAITANIS 300
 QY 301 IGAFLLIYASYALAFWYGTSLVLSSEYSIGQVLTVFFSVLIGAFSIGQASPSIEAFANA 360
 DB 301 IGAFLLIYASYALAFWYGTSLVLSSEYSIGQVLTVFFSVLIGAFSIGQASPSIEAFANA 360
 QY 361 RGAAYEIFKIIDNKPSIDSYSGHHPDNKGNLEFNHFSYPSRKEVKILKGLNLKVQ 420
 DB 361 RGAAYEIFKIIDNKPSIDSYSGHHPDNKGNLEFNHFSYPSRKEVKILKGLNLKVQ 420
 QY 421 SGOTVALVGNCGCKSTTVOLMORLYDPTDGMVCDIGQDITINVRLREITGVVSOEPV 480
 DB 421 SGOTVALVGNCGCKSTTVOLMORLYDPTDGMVCDIGQDITINVRLREITGVVSOEPV 480
 QY 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDLTVGERGAQLSGQKOR 540
 DB 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDLTVGERGAQLSGQKOR 540
 QY 541 IATARALVRNPKILLDEATSDLTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
 DB 541 IATARALVRNPKILLDEATSDLTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
 QY 601 AGFDDGVIVEKGNHDELMKEGIYFKLYVTMOTRGNIELENATGESKSDALENSPKDS 660
 DB 601 AGFDDGVIVEKGNHDELMKEGIYFKLYVTMOTRGNIELENATGESKSDALENSPKDS 660
 QY 661 GSSLIKRRSTRRSIHAPQGDRLGTKEDLNENVPVSWRILKLNSTWMPVVGIFCA 720
 DB 661 GSSLIKRRSTRRSIHAPQGDRLGTKEDLNENVPVSWRILKLNSTWMPVVGIFCA 720
 QY 721 IINGLOPAPSIIFRSRIIGFTREDEPETKRONSMFSLFLVLGIISFITFFLQGFTEG 780
 DB 721 IINGLOPAPSIIFRSRIIGFTREDEPETKRONSMFSLFLVLGIISFITFFLQGFTEG 780
 QY 781 KAGEIILTKRLRYMVFMSRLQDVSFDDPKNTTGALTTLANDAAQVKAIGSLAVITQ 840
 DB 781 KAGEIILTKRLRYMVFMSRLQDVSFDDPKNTTGALTTLANDAAQVKAIGSLAVITQ 840
 QY 841 NIANLGTIIISLIYQWLTLLALLAIVPIIATAGVVMKMLSQALKDKELEGAGKIAT 900
 DB 841 NIANLGTIIISLIYQWLTLLALLAIVPIIATAGVVMKMLSQALKDKELEGAGKIAT 900
 QY 901 EATENRTVVSILTREQKFYMAQSLQVPRNSLRKAHIFGVSFSTQAMMTFSTAGCPR 960
 DB 901 EATENRTVVSILTREQKFYMAQSLQVPRNSLRKAHIFGVSFSTQAMMTFSTAGCPR 960
 QY 961 FGAYLVANEFMNFODVLLVFSALVFGAMAVGVSSPAPDYAKAKYSAAHVIMTIEKSPIL 1020
 DB 961 FGAYLVANEFMNFODVLLVFSALVFGAMAVGVSSPAPDYAKAKYSAAHVIMTIEKSPIL 1020
 QY 1021 DSYSPHGLKPNTLEGNVTFNEVFNTPRPDIPVQLGSLVKKGTALVGVSSGCKST 1080
 DB 1021 DSYSPHGLKPNTLEGNVTFNEVFNTPRPDIPVQLGSLVKKGTALVGVSSGCKST 1080
 QY 1081 VVOLLERFDPLAGSVLIDGKEIKHLNVOWLAHIGIVSQEPIILDCSIAENIAYGDNRSR 1140
 DB 1081 VVOLLERFDPLAGSVLIDGKEIKHLNVOWLAHIGIVSQEPIILDCSIAENIAYGDNRSR 1140
 QY 1141 VVSHEEIMOAAKEANTHHFIETLPEKYNTRVGDGKQTLSSGOKORTATARALVRPHILL 1200
 DB 1141 VVSHEEIMOAAKEANTHHFIETLPEKYNTRVGDGKQTLSSGOKORTATARALVRPHILL 1200
 QY 1201 LDEATSDLTSEKVVQEAALDKAREGRTCIVIAHRLSTTIONADLIIVFQNGKVKHEGTHQ 1260
 DB 1201 LDEATSDLTSEKVVQEAALDKAREGRTCIVIAHRLSTTIONADLIIVFQNGKVKHEGTHQ 1260
 QY 1261 QLLAOKGIYFSVMVSOAGAKR 1281
 DB 1261 QLLAOKGIYFSVMVSOAGAKR 1281


```
RESULT 6
ID AAE00304 standard; Protein; 1280 AA.
XX AC AAE00304;
XX DT 13-JUN-2001 (first entry)
XX DE Dog P-glycoprotein (PGP) #2.
XX KW Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
XX KW drug bioavailability; transgenic animal; genetic model.
XX OS Canis familiaris.
XX PN WO200123540-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26767.
XX PR 28-SEP-1999; 99US-0156510.
XX PA (GENT-) GENTEST CORP.
XX PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX DR WPI: 2001-235373/24.
XX DR N-PSDB; AAD03489.
XX PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX PT for determining the bioavailability of drugs and for screening for dog
XX PT PGP inhibitors -
XX PS Example 2; Page 72-75; 11lpp; English.
XX CC The invention relates to dog P-glycoprotein (PGP) also referred
XX CC as multidrug transporter (MDR1) and nucleic acids encoding them.
XX CC The invention also includes fragments and biologically functional
XX CC variants of dog P-glycoprotein. PGP and their nucleic acids are
XX CC useful for determining the bioavailability of drugs and for
XX CC screening PGP inhibitors. They are useful for the diagnosis and
XX CC treatment of conditions characterised by PGP activity, by
XX CC reducing or increasing PGP activity in a cell. PGP nucleic acids
XX CC are used as oligonucleotide probes. Complements of PGP nucleic
XX CC acids are useful as antisense oligonucleotides, to induce a PGP
XX CC 'knockout' phenotype. They are used to prepare a non-human
XX CC transgenic animal, which are valuable as genetic models for
XX CC human diseases.
XX CC The present sequence is dog P-glycoprotein (PGP). The
XX CC PGP enzyme functions as an efflux pump exporting small molecules
XX CC across the cell membrane. This enzyme is a member of the ABC
XX CC transporter family.
XX SQ Sequence 1280 AA;

Query Match 99.3%; Score 6425.5; DB 22; Length 1280;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1273; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 1 MDPEGGRKGSAAENFWMGKKKKKKKKPTVSTFAMFRYSNWLDRMLVGTMAAII 60
DB 1 MDPEGGRKGSAAENFWMGKKK-KKKKKKKPTVSTFAMFRYSNWLDRMLVGTMAAII 59
QY 61 HGAALPLMLVFGNMTDSFANAGISRNKTPFVJINESITNNTQHFNHLEEMTTAYYY 120
DB 60 HGAALPLMLVFGNMTDSFANAGISRNKTPFVJINESITNNTQHFNHLEEMTTAYYY 119
QY 121 SGIGAGVLVAAYIQVSWFLAAGROILKIRKQFFHAIMRQIEGWFDVHDVGNLRLTDD 180
DB 120 SGIGAGVLVAAYIQVSWFLAAGROILKIRKQFFHAIMRQIEGWFDVHDVGNLRLTDD 179
QY 181 VSKINEGIDKGMFTQSIATFTFTGTVGTRGWKLTVLILATSPVLGSAIWAAILSS 240

Db 180 VSKINEGIDKGMFTQSIATFTFTGTVGTRGWKLTVLILATSPVLGSAIWAAILSS 239
QY 241 FTDKELLAYAKAGAAVEVLAAIIRTVIAFGGQKKELERYKNLEAEKIGIKKAITANIS 300
Db 240 FTDKELLAYAKAGAAVEVLAAIIRTVIAFGGQKKELERYKNLEAEKIGIKKAITANIS 299
QY 301 IGAAFLLIYASYALAFWYGTSLVLSSEYSIGQVLTFFSVLIGAFSIGQASPSIEAFANA 360
Db 300 IGAAFLLIYASYALAFWYGTSLVLSSEYTIGQVLTFFSVLIGAFSIGQASPSIEAFANA 359
QY 361 RGAAYEIFKIIDNKPSIDSYKSGHKPDNIKGNLEPKNFHFSYPSKVEKILKGLNLKVQ 420
Db 360 RGAAYEIFKIIDNKPSIDSYKSGHKPDNIKGNLEPKNFHFSYPSKVEKILKGLNLKVQ 419
QY 421 SGTVALVNSGCGKSTTVOLMORLYDPTDGMVCIDGQDIRTINVRLHREITGVVSQEPV 480
Db 420 SGTVALVNSGCGKSTTVOLMORLYDPTDGMVCIDGQDIRTINVRLHREITGVVSQEPV 479
QY 481 LFATTIAENIRYGRNVTWDEIEKAVKEANAYDFIMKLPNKFTDLVGERGALSGQKOR 540
Db 480 LFATTIAENIRYGRNVTWDEIEKAVKEANAYDFIMKLPNKFTDLVGERGALSGQKOR 539
QY 541 IATARALVRNPKILLDEATSAIDTSEAVVOVALDKARKGRTTIVIAHRLSTVRNADVI 600
Db 540 IATARALVRNPKILLDEATSAIDTSEAVVOVALDKARKGRTTIVIAHRLSTVRNADVI 599
QY 601 AGPDDGVIVEKGNHDELMKEKGIYFKLVMTQTRGNEIELEENATGESKSDALEMSPKDS 660
Db 600 AGPDDGVIVEKGNHDELMKEKGIYFKLVMTQTRGNEIELEENATGESKSDALEMSPKDS 659
QY 661 GSSLIKRRSTRRSIHAPOGQDRKLTGKEDLNENVPVSWRILKLNSTWEPVYVIGFICA 720
Db 660 GSSLIKRRSTRRSIHAPOGQDRKLTGKEDLNENVPVSWRILKLNSTWEPVYVIGFICA 719
QY 721 IINGGLQPAFSIIFRSIIIGIFTRDEDPETKRONSMFVLFLVLGIISITITFFLOGFTFG 780
Db 720 IINGGLQPAFSIIFRSIIIGIFTRDEDPETKRONSMFVLFLVLGIISITITFFLOGFTFG 779
QY 781 KAGEILTKRLRYMFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKAIGSLRAVITQ 840
Db 780 KAGEILTKRLRYMFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKAIGSLRAVITQ 839
QY 841 NTANLGTGIIISLIYGMQTLTLLAIPTIATAGVVEKMLSSQALKDKKELEGAGKAT 900
Db 840 NTANLGTGIIISLIYGMQTLTLLAIPTIATAGVVEKMLSSQALKDKKELEGAGKAT 899
QY 901 EATENFRTVYSLTREKQFEYMYAQSLQVYRNSLRKAHIFGVSFSTITQAMTYFSYAGCFR 960
Db 900 EATENFRTVYSLTREKQFEYMYAQSLQVYRNSLRKAHIFGVSFSTITQAMTYFSYAGCFR 959
QY 961 FGAYLVANEFMNFQDVLVFSIAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPLI 1020
Db 960 FGAYLVANEFMNFQDVLVFSIAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPLI 1019
QY 1021 DSYSPHGLKPNLTLEGNTFNEVVENYPTRPDIPVLQGLSLEVKKGQTLALVSSCGCKST 1080
Db 1020 DSYSPHGLKPNLTLEGNTFNEVVENYPTRPDIPVLQGLSLEVKKGQTLALVSSCGCKST 1079
QY 1081 VYQLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHLGIVSQEPILFDCSTAENIAYGDNRS 1140
Db 1080 VYQLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHLGIVSQEPILFDCSTAENIAYGDNRS 1139
QY 1141 VYSHEIEMQAQAEANHHFIETLPKYNTRVGDKGTQISGGQKQRIATARALVRPHILL 1200
Db 1140 VYSHEIEMQAQAEANHHFIETLPKYNTRVGDKGTQISGGQKQRIATARALVRPHILL 1199
QY 1201 LDEATSAIDTSEKVVQALDKAREGRTCVIAHRLSTIQNADLIVFPQNKVKEHGTHQ 1260
Db 1200 LDEATSAIDTSEKVVQALDKAREGRTCVIAHRLSTIQNADLIVFPQNKVKEHGTHQ 1259
QY 1261 QLLAQKGIYFMSVVOAGAKR 1281
XXXXXXXXXXXXXXXXXXXX
```

Db 1260 QLLAQKGIYFSMISVQAGAKR 1280

RESULT 7

AAW44073

ID AAW44073 standard; protein; 1280 AA.

XX

XX

AC AAW44073;

XX

DT 26-JUN-1998 (first entry)

XX

DE Human multidrug resistance P-glycoprotein MDRI.

XX

XX Human; multidrug resistance P-glycoprotein; MDRI; prokaryotic homologue.

KW

XX

OS Homo sapiens.

XX

FH Key

Region

1..640

/note= "MDRI-N from Fig 1"

FT

Region

641..1280

/note= "MDRI-C from Fig 1"

FT

XX

XX

PN W09740160-A1.

XX

XX

PD 30-OCT-1997.

XX

XX

PF 24-APR-1997; 97WO-N000216.

XX

XX

PR 24-APR-1996; 96EP-0201094.

XX

XX

PA (UYGR-) RIJKSUNIV GRONINGEN.

XX

XX

PI Bolhuis H, Konings WN, Van Veen HW, Venema K;

XX

XX

DR WPI; 1997-535844/49.

XX

DR

XX

PT Prokaryotic homologue of human multiple drug resistance protein -

XX

PT used to screen for compounds that inhibit, or avoid, drug resistance

XX

XX

XX

PS Claim 10; Fig 1; 35pp; English.

XX

XX

CC The present invention describes a recombinant or isolated nucleic acid

CC (1), derived from a prokaryotic gene, which encodes at least a specific

CC and/or functional part of a transporter protein (TP), or its

CC derivatives, which has functional and/or structural similarity with the

CC P-glycoprotein (PG) encoded by the human multidrug resistance

CC (MDR)1 gene. The present sequence represents the human MDRI protein,

CC derived from MDRI-N and MDRI-C as shown in the specification in

CC figure 1. (1) is used to express recombinant proteins; its fragments

CC are also useful as probes and primers for detection and amplification

CC of related DNA. The protein produced, or cells expressing them, are

CC used to determine if substances can inhibit, or avoid, MDR proteins,

CC and in a screening method for identifying compounds that inhibit

CC transport of cytotoxic substances from cells. Also, cells with a

CC transmembrane protein, especially where expressed from (1), can provide

CC (additional) MDR, particularly for use as a model system to study

CC mechanisms of action of PG.

XX

SQ

Sequence 1280 AA;

Query Match 90.6%; Score 5866; DB 18; Length 1280;

Best Local Similarity 90.8%; Pred. No. 0;

Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;

QY 1 MDPEGGKRGSA-EKNFWKWKSKKKEKKKPTVSTFAMFRYSNWLDRILYVGTMAAI 59

Db 1 MDLEGRNGAKKKNFKLNKS-EKDKKPKTYSVFSMFRYSNWLDRILYVGTMAAI 59

QY 60 IHGAALPLMLVFGNMTDSFANAGISRNKTFPVVINESITNNTOHFINHLEEMTYAY 119

Db 60 IHGAGLPLMLVFGWTDIFANAG-NLEDLMSNITNRSINDNGFFMN-LEEDMTYAY 117

QY 120 YSGIGAGVLVAAYIQVSWFCLAAAGRQILKIRKQFFHAIMRQETGWFVDVHVGELNTRLTD 179

Db 118 YSGIGAGVLVAAYIQVSWFCLAAAGRQILKIRKQFFHAIMRQETGWFVDVHVGELNTRLTD 177

QY 180 DYSKINEGIGDKIGMFFQSTATFTFTGTRGKWLTLVLAISPVGLSAAIWAAILKS 239

Db 178 DYSKINEGIGDKIGMFFQSMATFTFTGTRGKWLTLVLAISPVGLSAAIWAAILKS 237

QY 240 SFTDKELLAYAKAGAVAEVLAAIRTVIAFGGQKKELERYNKNLEAKRIGIKKAITANI 299

Db 238 SFTDKELLAYAKAGAVAEVLAAIRTVIAFGGQKKELERYNKNLEAKRIGIKKAITANI 297

QY 300 SIGAAFLIYASYALAPWYGTSLVLSSEYSIGQVLTVPFFSVLIGAFSIGQASPSIEAFAN 359

Db 298 SIGAAFLIYASYALAPWYGTSLVLSSEYSIGQVLTVPFFSVLIGAFSIGQASPSIEAFAN 357

QY 360 ARGAAEYELFKIIDNKPSTDSYKSGHKPDNIKGNLEKFNHFSYPSRKEVKILGLNLKV 419

Db 358 ARGAAEYELFKIIDNKPSTDSYKSGHKPDNIKGNLEKFNHFSYPSRKEVKILGLNLKV 417

QY 420 QSGQTVLGVNSGCGKSTTVQLMORLYDPTDMVCIDGQDITINVRHLREITGVWSQEP 479

Db 418 QSGQTVLGVNSGCGKSTTVQLMORLYDPTDMVCIDGQDITINVRHLREITGVWSQEP 477

QY 480 VLFATTIAENTRYGRENVTMDIEKAVKANAYDFIMKLPNKFTLVGERGAQLSGGQK 539

Db 478 VLFATTIAENTRYGRENVTMDIEKAVKANAYDFIMKLPNKFTLVGERGAQLSGGQK 537

QY 540 RIATARALVRNPKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSVRNADV 599

Db 538 RIATARALVRNPKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSVRNADV 597

QY 600 IAGDDGVIVEKGNHDELMKEGIVFKLVMTQTRGNEIELEENATGESKSDALEMSPKD 659

Db 598 IAGDDGVIVEKGNHDELMKEGIVFKLVMTQTRGNEIELEENATGESKSDALEMSPKD 657

QY 660 SGSSLIKRRSTRRSIHAPQGDQRKLGTKEDLNENVPVFWRIILKLNSTWEPYVVGIFC 719

Db 658 SRSLIRKRRSTRRSVRSQAQDRKLSKEALDESIPPVFWRIILKLNSTWEPYVVGIFC 717

QY 720 ALINGGLQPAFSITFSRIIGITRDEDPETKRONSMFSLVLVLGIISFTFFLQGTFF 779

Db 718 ALINGGLQPAFSITFSRIIGITRDEDPETKRONSMFSLVLVLGIISFTFFLQGTFF 777

QY 780 GKAGEILTKRLRYMVRSMRLQDVSFDDPKNTTGALTTRLANDAAQVKGAGISRLAVIT 839

Db 778 GKAGEILTKRLRYMVRSMRLQDVSFDDPKNTTGALTTRLANDAAQVKGAGISRLAVIT 837

QY 840 QNIANLGTGIIISLIYHQWLLLLAIVPIITAIAGVEMKMLSGOALKDKKELEGAGKIA 899

Db 838 QNIANLGTGIIISLIYHQWLLLLAIVPIITAIAGVEMKMLSGOALKDKKELEGAGKIA 897

QY 900 TEAIENFTVSLTREQKFEYMYAQSLQVYRNSLRKAHIFGVFSFISITQAMMYFSYAGCF 959

Db 898 TEAIENFTVSLTREQKFEYMYAQSLQVYRNSLRKAHIFGVFSFISITQAMMYFSYAGCF 957

QY 960 RFGAYLVANEFMNFODVLLVFSIAIVFGAMAVQGVSSFPADYAKAKVSAHVIMIEKSP 1019

Db 958 RFGAYLVANEFMNFODVLLVFSIAIVFGAMAVQGVSSFPADYAKAKVSAHVIMIEKSP 1017

QY 1020 IDSYSPHGLKPNLTLEGNTFNEVFNPTRPDIPLVQLGSLSEVKKGOTLALVSGSGGCKS 1079

Db 1018 IDSYSPHGLKPNLTLEGNTFNEVFNPTRPDIPLVQLGSLSEVKKGOTLALVSGSGGCKS 1077

QY 1080 TVVQLLEFYDPLAGSLVDGKEIKHLNVQWLRHLGLIVSQEPILFDCSIAENIAYGDN 1139

Db 1078 TVVQLLEFYDPLAGSLVDGKEIKHLNVQWLRHLGLIVSQEPILFDCSIAENIAYGDN 1137

QY 1140 RVVSHEEIMQAAKEANIHHFTETLPEKYNTRYGVGDKGTOLSGGQKQRTIAIALRVOPHIL 1199

Db 1138 RVVSHEEIMQAAKEANIHHFTETLPEKYNTRYGVGDKGTOLSGGQKQRTIAIALRVOPHIL 1197

QY 1200 LLDEATSAIDTSEKVKVQEAIDKAREGTCIVIAHRLSTIQNADLIVVVFQNGKVEHGT 1259

Db 1018 IDSYSYGLKPNTLEGNTFFNEVFNPTRLDIPVLQGLSLEVKKGTALVSSGCGKS 1077
QY 1080 TVVOLLERYDPLAGSVLDGKEIKHLNQLRAHLGIVSQPILFDCSIAENIAYGNS 1139
Db 1078 TVVOLLERYDPLAGVLDGKEIKQLNVQLRAHLGIVSQPILFDCSISENIAYGNS 1137
QY 1140 RVYSHEEIMQAAKEANIHFIEPLPEKYNTRVGDKGTQLSGGQKQRIATARALVRQPHIL 1199
Db 1138 RVYSQEEIVRAAKEANIHFIESLPNKYSTRVGDKGTQLSGGQKQRIATARALVRQPHIL 1197
QY 1200 LDEATSAIDTSEKVVQBALDKAREGRTCIIVIAHRLSTIQNADLIVFQNGKVKHEGTH 1259
Db 1198 LDEATSAIDTSEKVVQBALDKAREGRTCIIVIAHRLSTIQNADLIVFQNGRVKHEGTH 1257
QY 1260 QLLAOKGIVFMSVQAGAKR 1281
Db 1258 QLLAOKGIVFMSVQAGAKR 1279.
RESULT 11
AAB81065
ID AAB81065 standard; Protein; 1283 AA.
XX AAB81065;
AC AAB81065;
XX
DT 25-JUN-2001 (first entry)
XX
DE Cynomologous monkey P-glycoprotein variant 2.
DE
DE Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL1;
KW efflux pump.
KW
KW
OS Macaca fascicularis.
FH
FH Key Location/Qualifiers
FT Misc-difference 93..95
FT /note= "An additional 3 amino acids are present compared
FT to PGP variant AAB81064"
XX
XX WO200123565-A1.
XX
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26592.
XX
XX 28-SEP-1999; 99US-0156921.
PR 12-OCT-1999; 99US-0158818.
XX
XX (GENT-) GENTEST CORP.
XX
XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
PI WPI; 2001-316136/33.
XX N-PSDB; AAF86128.
DR
DR
XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell -
XX
XX
PS Claim 9; Page 65-68; 84pp; English.
XX
CC This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDRL1 is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents the

CC cynomologous monkey P-glycoprotein variant 2. The protein has an
CC additional 3 amino acids when compared to PGP variant 1 (AAB81065).
XX
SQ Sequence 1283 AA;
Query Match 90.5%; Score 5859.5; DB 22; Length 1283;
Best Local Similarity 90.4%; Pred. No. 0;
Matches 1165; Conservative 57; Mismatches 52; Indels 15; Gaps 5;
QY 1 MDPEGKRGSAE-KNFKMGKKKKKKPTVSTFAMFRYSNWLDRMLVGVMAAI 59
Db 1 MDLEGRNGGAENKFFLNKSK-KDKKPKPTVSVFMSFRYSNWLDRMLVGVMAAI 59
QY 60 IHGAALPLMLVFGNMTDSFANAG-----ISRNKTFPVIINESITNTQTHFINHLEEE 112
Db 60 IHGAGLPLMLVFGMDTDTFANAGNLGDLGALLFNNT-----NSSNITDTPVPMN-LEED 113
QY 113 MTTAYYYSGIGAGVIAVAAIYQVSWFCLAAAGROILKIRKOFFHAIMRQEIQWDFDHDVGE 172
Db 114 MTRYAYYSIGAGVIAVAAIYQVSWFCLAAAGROILKIRKOFFHAIMRQEIQWDFDHDVGE 173
QY 173 LNRLTDDVSKINEGIGDKIGMFFQSIATFTFTGIVGTRGWKLTILVLAISPVLGSA 232
Db 174 LNRLTDDVSKINEGIGDKIGMFFQSMATFTFTGIVGTRGWKLTILVLAISPVLGSA 233
QY 233 IWAKILSFTDKELLAYAKAGAAEVLAAIRTVIAFGGOKKELERYNNKLEAKGIGIK 292
Db 234 VWAKILSFTDKELLAYAKAGAAEVLAAIRTVIAFGGOKKELERYNNKLEAKGIGIK 293
QY 293 KAITANTISIGAAFLIYASALAFWYGTTLVLSKEYSIGQVLTFFVSVLGAFSVGOASP 352
Db 294 KAITANTISIGAAFLIYASALAFWYGTTLVLSKEYSIGQVLTFFVSVLGAFSVGOASP 353
QY 353 SIEAFANARGAAEYIFKLIIDNKPSIDSYSGHGPDKIKGNLEPKNVHFSYPSRKEVKIL 412
Db 354 SIEAFANARGAAEYIFKLIIDNKPSIDSYSGHGPDKIKGNLEPNVHFSYPSRKEVKIL 413
QY 413 KGLNLKVSQGTVALVNSGCGKSTTVQLMORLYDPTDGMVCIDGDIRINVRHLREIT 472
Db 414 KGLNLKVSQGTVALVNSGCGKSTTVQLMORLYDPTDGMVSDGQDIRINVRFLREII 473
QY 473 GVSQEPVLFPATTAEINIRYGRNVMTDEIKAVKEANAYDFIMKLPNKDFTLVGERGAQ 532
Db 474 GVSQEPVLFPATTAEINIRYGRNVMTDEIKAVKEANAYDFIMKLPNKDFTLVGERGAQ 533
QY 533 LSGGQKQRIATARALVRNPVKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLS 592
Db 534 LSGGQKQRIATARALVRNPVKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLS 593
QY 593 TVRNADVIAGFDDGVIVVEKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGESKESDA 652
Db 594 TVRNADVIAGFDDGVIVVEKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGESKESIDT 653
QY 653 LEMSPKDSGLIKRRSTRRSIHAPOGQDRKLGTKEDLINEPVPPVFWIRILKLNSTEWPY 712
Db 654 LEMSHDGSGLSIKRRSTRRSVRGSGQDRKLSRKEALDESIPPVFWIRILKLNTEMPY 713
QY 713 FVVGIFCAIINGGLQPAFSAIIFSRRIIGFTTRDEPDKRONSMFSLFLVLGIISITTF 772
Db 714 FVVGIFCAIINGGLQPAFAVIFSKIIIGFTTRNDDEATKRONSLFSLFLVLGIISITTF 773
QY 773 FLOQFTFGKAGEILTTLRLYVMFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIG 832
Db 774 FLOQFTFGKAGEILTTLRLYVMFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIG 833
QY 833 SRLAVITONIANLGTGIIISLIYQWLTLLLAIVPIIATAGVVEKMLSKQALKDKKEL 892
Db 834 SRLAVITONIANLGTGIIISLIYQWLTLLLAIVPIIATAGVVEKMLSKQALKDKKEL 893
QY 893 EGAGKIATEAIENTFRVTVSLTREQKFEYMYAQSLQVPRNSLRKAHIFGVFSFTQAMMY 952
Db 894 EGAGKIATEAIENTFRVTVSLTREQKFEHMYDQSLQVPRNSLRKAHIFGVFSFTQAMMY 953

Db 658 SRSSLIKRSTRSRVSGSQADQKRLSTKALDESIPPVSFWRIMKLNLTENPWFYVGVFC 717
Qy 720 AINGGLQAPAFSIISRIIGITREDEPETKRONSMESVLFVLGLIISFITEFLOGTFE 779
Db 718 AINGGLQAPAFSIISKIIGVFTRIDDPETKRONSLFLSLFLALGIIISFITEFLOGTFE 777
Qy 780 KGAGELLKRLRYMFRSMRLQDVSWFDDPKNTTGALTTRLANDAAQVKGAGISRLAVIT 839
Db 778 KGAGELLKRLRYMFRSMRLQDVSWFDDPKNTTGALTTRLANDAAQVKGAGISRLAVIT 837
Qy 840 QNIANLGTGIISLYGWOLTLALLAIPIIAIAGVEMKMLSGQALKDKKLEBAGAKIA 899
Db 838 QNIANLGTGIISLYFIYGMOLTLALLAIPIIAIAGVEMKMLSGQALKDKKLEBAGAKIA 897
Qy 900 TEAIENFRVVSITREOKFEYMYAQSLOWPYRNSLRKAHIFGVSPSITQAMMYFSYAGCF 959
Db 898 TEAIENFRVVSITREOKFEYMYAQSLOWPYRNSLRKAHIFGVSPSITQAMMYFSYAGCF 957
Qy 960 RFGAYLVANEFMNFQDVLVFSIAIVFGAMAVGVQVSSFAPDYAKAKVSAAHVIMIEKSPL 1019
Db 958 RFGAYLVANEFMNFQDVLVFSIAIVFGAMAVGVQVSSFAPDYAKAKVSAAHVIMIEKSPL 1017
Qy 1020 IDSYPHGLKPNLTENVTNEVNYPRPDIPVLQGLSLEVKKGQTLALVSGSGCGKS 1079
Db 1018 IDSYSTEGLPNTLEGNVTFGEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSGSGCGKS 1077
Qy 1080 TVVOLLERFYDPLAGSVLDGKEIKHLNQLWRAHLGIVSQEPILFDCSIAENIAYGNS 1139
Db 1078 TVVOLLERFYDPLAGSVLDGKEIKHLNQLWRAHLGIVSQEPILFDCSIAENIAYGNS 1137
Qy 1140 RVVSHEEIMQAAKEANIHFIEPLPEKYNTRYGDKGTQLSGGQKQRIARALVRQPHIL 1199
Db 1138 RVVSQEEIVRAAKEANIHFIESLPNKYSTKVGDKGTLSGGQKQRIARALVRQPHIL 1197
Qy 1200 LDEATSAIDTESEKVVQALDOKAREGRTCVIAHRLSTIQNADLIVVFQNGKVKHEGTH 1259
Db 1198 LDEATSAIDTESEKVVQALDOKAREGRTCVIAHRLSTIQNADLIVVFQNGKVKHEGTH 1257
Qy 1260 QLLAOKGIFYSMVSVQAGAKR 1281
Db 1258 QLLAOKGIFYSMVSVQAGTKR 1279
RESULT 13
AAB81066
ID AAB81066 standard; Protein; 1280 AA.
XX AAB81066;
AC AAB81066;
DT 25-JUN-2001 (first entry)
XX Human P-glycoprotein SEQ ID 5.
DE Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;
XX efflux pump; human.
KW Homo sapiens.
XX OS
XX WO200123565-A1.
PN
XX
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26592.
XX
PR 28-SEP-1999; 99US-0156921.
PR 12-OCT-1999; 99US-0158818.
XX
XX (GENT-) GENTEST CORP.
PA
XX
XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
PI WPI; 2001-316136/33.
XX
DR

XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell -
XX
PS Claim 6; Page 68-71; 84pp; English.
XX
XX This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDR1 is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents a human
CC P-glycoprotein.
XX
SQ Sequence 1280 AA;
Query Match 90.5%; Score 5857; DB 22; Length 1280;
Best Local Similarity 90.7%; Pred. No. 0;
Matches 1163; Conservative 55; Mismatches 60; Indels 4; Gaps 4;
Qy 1 MDPEGGRKGS- EKNFWKMGKSKKKEKPKTPVSTFAMFRYSNWLDRMLVGTMAAI 59
Db 1 MDLEGRNGGAKKKNFKLNKS- EKDKKEKPKTPVSFMSFRYSNWLDRMLVGTMAAI 59
Qy 60 IHGAALPLMLVFNMTDTSFANAGISRNKTPFPVVIINESITNTQHFINHLEETTYAYY 119
Db 60 IHGAGLPLMLVFGEMTDIFANAG- NLEDLMSNTNRSINDTGFNMN- LEEDMTRYAYY 117
Qy 120 YSGIGAGVLVAAVYQVSWFCLAGROQLKTKQFFHAIMRQEGWFDVHDVDELNRLTD 179
Db 118 YSGIGAGVLVAAVYQVSWFCLAGROQLKTKQFFHAIMRQEGWFDVHDVDELNRLTD 177
Qy 180 DVSKINEGIDKIGMFFQSIATFTFTGIVGTRGWLKTLVLILASPVLGSLAAWAKILS 239
Db 178 DVSKINEVIGDKIGMFFQSMATFTFTGIVGTRGWLKTLVLILASPVLGSLAAWAKILS 237
Qy 240 SFTDKELLAYAKAGAAVEVLAAITRTIAGGGOKKELERNKNLEAKGIGIKKAITANI 299
Db 238 SFTDKELLAYAKAGAAVEVLAAITRTIAGGGOKKELERNKNLEAKRIGIKKAITANI 297
Qy 300 SIGAFLLIYASYALAFWYGTSLVLSSEYSIGQVLTVPFVSVLIGAFSIGQASPSIEAFAN 359
Db 298 SIGAFLLIYASYALAFWYGTSLVLSSEYSIGQVLTVPFVSVLIGAFSIGQASPSIEAFAN 357
Qy 360 ARGAAYEIFKIIDNKPSIDSYSGSHKPDNIKGNLEFKNVHFSPSRKEVKILKGLNLKV 419
Db 358 ARGAAYEIFKIIDNKPSIDSYSGSHKPDNIKGNLEFKNVHFSPSRKEVKILKGLNLKV 417
Qy 420 QSGQTVALVGNSSGCKSTTVOLMORLDPTDGMVCIDGQDRTINVRHLREITGVVSOEP 479
Db 418 QSGQTVALVGNSSGCKSTTVOLMORLDPTDGMVSVSDGQDRTINVRHLREITGVVSOEP 477
Qy 480 VLFATTIAENIRYGRNVNTMDEIEKAVKEANAYDFIMKLPKFDTLVGERGAQLSGGQKQ 539
Db 478 VLFATTIAENIRYGRNVNTMDEIEKAVKEANAYDFIMKLPKFDTLVGERGAQLSGGQKQ 537
Qy 540 RIATARALVRNPKIILLDEATSAIDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 599
Db 538 RIATARALVRNPKIILLDEATSAIDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 597
Qy 600 IAGFDDGVIVEKGNHDELMKEGIYFKLVMTQTRNETELEENATGESKSDALEMSPKD 659
Db 598 IAGFDDGVIVEKGNHDELMKEGIYFKLVMTQTRNETELEENATGESKSDALEMSPKD 657
Qy 660 SGSSLIKRSTRSRSIHAPQGDQKRLGKEDLNENVPVSWFRILKLNSTENPWFYVGVFC 719
Db 658 SGSSLIKRSTRSRSIHAPQGDQKRLGKEDLNENVPVSWFRILKLNSTENPWFYVGVFC 719

Db 658 SRSSLIRKRRSRVSGAQRKLSKTEALDESIPPVFWKIMKLNLTENWPYFVVGVC 717
QY 720 AINGLOPAPSLFSLRIGIFTRDEDEPETHRONSNFSLFLVGLIISFITFFLOGTF 779
Db 718 AINGLOPAPSLFSLRIGIFTRIDDPETHRONSNFSLFLVGLIISFITFFLOGTF 777
QY 780 KGAGELFKRLRYMVRFRSMLRQDVSFDDPKNTTCALTRLANDAAQVKGAGISRLAVIT 839
Db 778 KGAGELFKRLRYMVRFRSMLRQDVSFDDPKNTTCALTRLANDAAQVKGAGISRLAVIT 837
QY 840 QNIANLGTGIIISLYGQHLPLLLAIVPIITAIAGVEMKMLSGQALKDKKELSGAKIA 899
Db 838 QNIANLGTGIIISLYGQHLPLLLAIVPIITAIAGVEMKMLSGQALKDKKELSGAKIA 897
QY 900 TEATENFTVYSLTRQOKFEYMAQSLQVPYRNSLRKAHIFGVSFSTQAMWYSYAGCF 959
Db 898 TEATENFTVYSLTRQOKFEYMAQSLQVPYRNSLRKAHIFGVSFSTQAMWYSYAGCF 957
QY 960 RFGAYLVANEFMFDVLLVFSIAVFGAMAVGOVSSFPADYAKARVAAHVIMIIEKSPL 1019
Db 958 RFGAYLVANEFMFDVLLVFSIAVFGAMAVGOVSSFPADYAKARVAAHVIMIIEKSPL 1017
QY 1020 IDSYSPLKNTLEGNTFNEVFNPTPRDIPVLOGLSLEVKKGTALVSGSGCGKS 1079
Db 1018 IDSYSPLKNTLEGNTFNEVFNPTPRDIPVLOGLSLEVKKGTALVSGSGCGKS 1077
QY 1080 TVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHGLGVSOEPLDFCSTAENIAYGDNS 1139
Db 1078 TVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHGLGVSOEPLDFCSTAENIAYGDNS 1137
QY 1140 RVSHIEIMAAKANIHHFETLPEKYNTRVGDGKQLSGGQKQRIATAIARALVRQPHIL 1199
Db 1138 RVSHIEIMAAKANIHHFETLPEKYNTRVGDGKQLSGGQKQRIATAIARALVRQPHIL 1197
QY 1200 LLDATSAIDTESKVVQVQALDKAREGTCIVIAHRLSTIQNADLIIVFONGKVKHEGTH 1259
Db 1198 LLDATSAIDTESKVVQVQALDKAREGTCIVIAHRLSTIQNADLIIVFONGKVKHEGTH 1257
QY 1260 QOLLAQKGIYFMSVVOAGAKR 1281
Db 1258 QOLLAQKGIYFMSVVOAGAKR 1279
RESULT 14
AAE00306
ID AAE00306 standard; Protein; 1280 AA.
AC AAE00306;
XX
XX 13-JUN-2001 (first entry)
DT Human P-glycoprotein (PGP) #1.
DE
DE Human P-glycoprotein (PGP) #1.
KW Human; P-glycoprotein; PGP; multidrug transporter; MDRI;
KW drug bioavailability; transgenic animal; genetic model.
XX
XX Homo sapiens.
XX
XX W0200123540-A2.
PN
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26767.
PF
XX
PR 28-SEP-1999; 99US-0156510.
XX
XX (GENT-) GENTEST CORP.
PA
XX
XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
PI
XX
XX WPI; 2001-235373/24.
DR
XX
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful

PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
PS Claim 16; Page 78-80; 11pp; English.
XX
CC The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDRI) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is human P-glycoprotein (PGP). The
CC PGP enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.
XX

SQ Sequence 1280 AA;

Query Match 90.5%; Score 5857; DB 22; Length 1280;

Best Local Similarity 90.7%; Pred. No. 0;
Matches 1163; Conservative 55; Mismatches 60; Indels 4; Gaps 4;

QY 1 MDPEGGKGS-A-EKNFWKMKSKKKEKPTVTFAMFRYSNWLDRYMLVGTMAAI 59

Db 1 MDLEGRNGGAKKNFKLNKS-EKDKEKPTVSVFMSNWLDRYMLVGTMAAI 59

QY 60 IHGAALPLMLVPGNMTDSFANAGISRNTFPVINESITNTQHFHINLEEMTYAY 119

Db 60 IHGAGLPLMLVPGNMTDSFANAG-NLEDMSNITNRSINDTGFNM-LEEDTRYAY 117

QY 120 YSGIGAGVLAAYIQVSWCLAGROILKIQOFHAIMRQETGWFVDHVGELNRLTD 179

Db 118 YSGIGAGVLAAYIQVSWCLAGROILKIQOFHAIMRQETGWFVDHVGELNRLTD 177

QY 180 DVSKINEGIDKIGMFFQSIATFTFTGIVGTRGWKLTILVLAISPVGLSAIAWAKILS 239

Db 178 DVSKINEGIDKIGMFFQSIATFTFTGIVGTRGWKLTILVLAISPVGLSAIAWAKILS 237

QY 240 SFTDKELLAYKAGAAVEVLAIRTIVIAFGQKKELERYKNLEAKGIGIKKAITANI 299

Db 238 SFTDKELLAYKAGAAVEVLAIRTIVIAFGQKKELERYKNLEAKGIGIKKAITANI 297

QY 300 SIGAAFLIYASVALAFWYGTSLVLSSEYSIGOVLTVPFVSIGQASPSFEAFAN 359

Db 298 SIGAAFLIYASVALAFWYGTSLVLSSEYSIGOVLTVPFVSIGQASPSFEAFAN 357

QY 360 ARGAAVEIFKIIDNKNPSIDSYSGHKKPNIKGNLEFKNVHFSPSRKEVTKLGNLKV 419

Db 358 ARGAAVEIFKIIDNKNPSIDSYSGHKKPNIKGNLEFKNVHFSPSRKEVTKLGNLKV 417

QY 420 QSGQTVLVGNSGCGKSTTVQMLQRLYDPTDGMVIGDQIDRTINVRHLREITGVVSE 479

Db 418 QSGQTVLVGNSGCGKSTTVQMLQRLYDPTDGMVIGDQIDRTINVRHLREITGVVSE 477

QY 480 VLFATTIAENIRYGRNVTMDIEKAVKANAYDFIMKLPNKFDTLVGERGQALSGGOK 539

Db 478 VLFATTIAENIRYGRNVTMDIEKAVKANAYDFIMKLPNKFDTLVGERGQALSGGOK 537

QY 540 RIAIARALVRNPKILLDEATSAIDTESAVVQVVALDKARKGRTTIVIAHRLSTVNRAD 599

Db 538 RIAIARALVRNPKILLDEATSAIDTESAVVQVVALDKARKGRTTIVIAHRLSTVNRAD 597

QY 600 IAGFDGIVVEKGNHDELMKEKGIYFKLVMTQTRGNETELENATGESKESDALBMSPKD 659

Db 598 IAGFDGIVVEKGNHDELMKEKGIYFKLVMTQTRGNETELENATGESKESDALBMSPKD 657

QY 660 SGSSLLKRSTRSRSHAPQODRKLGTREDNLNENPPVSWFRILKLNSTWPFVVGIFC 719
 DB 658 SRSSLLKRSTRSRVSGSOADRKLSTKEALDESIPPSWFRIMKLNLTWPFVVGVC 717
 QY 720 AINGLQPAFISIFRSIGITFREDPETRQRNSMFSVLVLGUISFTFFLQGTFF 779
 DB 718 AINGLQPAFAIFISKIIGVTRDDPETRQRNSLFLALGUISFTFFLQGTFF 777
 QY 780 GRAGEILRLRYMYFRSMRLQDVSWFDDPKNTTALTRLANDAAQVKGAGSLAVIT 839
 DB 778 GRAGEILTRLRYMYFRSMRLQDVSWFDDPKNTTALTRLANDAAQVKGAGSLAVIT 837
 QY 840 QNANLGTGIIISLYGWQLTLLLAIVPIIAIAGVWEMKMLSGQALKDKELEGAGKIA 899
 DB 838 QNANLGTGIIISLYGWQLTLLLAIVPIIAIAGVWEMKMLSGQALKDKELEGAGKIA 897
 QY 900 TEAIENFRVVSLSLTREQPEYMYAQSLOVPYRNSLRKAHIFGVSPSTQAMMYFYAGCF 959
 DB 898 TEAIENFRVVSLSLTREQPEYMYAQSLOVPYRNSLRKAHIFGVSPSTQAMMYFYAGCF 957
 QY 960 RGCAYLVANEFNFQDVLVLSAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMITEKSPL 1019
 DB 958 RGCAYLVANHLMSFEDVLVLSAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMITEKPL 1017
 QY 1020 IDSYSPLGLKPNTEGNTVFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS 1079
 DB 1018 IDSYSPEGLMPNTEGNTVFGEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS 1077
 QY 1080 TVVQLLERYDPLAGSVLDGKEIKHLNVQWIRAHLGIVSQBPILFDGCSIAENIAYGNS 1139
 DB 1078 TVVQLLERYDPLAGSVLDGKEIKHLNVQWIRAHLGIVSQBPILFDGCSIAENIAYGNS 1137
 QY 1140 RVVSHEEIMOAAKEANIHFIEITLPEKYNTVRGDKGTQSLGGQKORIAARALVRQPHIL 1199
 DB 1138 RVVSQEEIVRAAKEANIHFIESLPNKYSTKVGDKGTQSLGGQKORIAARALVRQPHIL 1197
 QY 1200 LLDATSAIDTESEKVVQALDKAREGRTCVIAHRLSTIQNADLIVVFQNGKVKHEGTH 1259
 DB 1198 LLDATSAIDTESEKVVQALDKAREGRTCVIAHRLSTIQNADLIVVFQNGKVKHEGTH 1257
 QY 1260 QLLAOKGIYFSMVSVQAGAKR 1281
 DB 1258 QLLAOKGIYFSMVSVQAGTKR 1279

RESULT 15
 AAP70452
 ID AAP70452 standard; Protein; 1280 AA.
 AC
 XX
 XX
 XX
 DT 21-MAY-1991 (first entry)
 XX
 DE Sequence encoded by human multi-drug resistance-1 (mdr1) cDNA
 DE from clones lambda-HDR10.5 and 104.
 XX
 KW Chemo-therapy resistant tumour cell; P-glycoprotein.
 XX Homo sapiens.
 OS
 PN W08705943-A.
 XX
 XX 08-OCT-1987.
 XX
 XX 26-MAR-1987; 87WO-US00758.
 XX
 PR 01-AUG-1986; 86US-0892575.
 PR 28-MAR-1986; 86US-0845610.
 XX
 XX (UNII) UNIV OF ILLINOIS.
 XX
 XX Roninson IB, Pastan IH, Gottesman MM;

DR WPI; 1987-291656/41.
 XX N-PSDB; AAN70752.
 XX
 PT DNA for multi-drug resistance in human cells - used to detect
 PT chemotherapy-resistant tumour cells and for producing
 PT polypeptide(s) for diagnosis and therapy
 XX
 PS Claim 4(a); Table 5, pp30-39; 61pp; English.
 XX
 CC The human multi-drug resistant KB carcinoma cell lines were used as
 CC the source of the mdr1 gene nucleic acid sequences (AAN70751). To
 CC obtain cDNA clones of the mdr1 gene (AAN70752), poly (A) and RNA was
 CC used. Analysis of the AA sequence presented in (AAN70752) indicates
 CC that the mdr1 gene product is likely to be a transmembrane protein.
 CC The presence of transmembrane domains and potential glycosylation
 CC sites is consistent with the mdr1 protein being related to the
 CC P-glycoprotein.
 XX
 SQ Sequence 1280 AA;
 Query Match 90.4%; Score 5854; DB 8; Length 1280;
 Best Local Similarity 90.6%; Pred. No. 0;
 Matches 1162; Conservative 56; Mismatches 60; Indels 4; Gaps 4;
 QY 1 MDPEGGRKGS-A-EKNFMGMKKKKKKKKKPTVSTFAMFRYSNWLDRILMLVGTMAAI 59
 DB 1 MDLEGDRNGGAKKNFFKLNKKS-EKDKKKKPTVSFMSFRYSNWLDRILMLVGTMAAI 59
 QY 60 IHGAALPLMLVFCNMVDSFANAGISRNKTFPPVITNESITNTQHFINHLEEMETVYAY 119
 DB 60 IHGAGLPLMLVFCNMVDSFANAG-NLEDLMSNTNRSDINDTGFFMN-LEEDWTRVAY 117
 QY 120 YSGIGAGVLAAYIQVSWFCLAGRQILKIRKQFFHAIMRQEIWFVDVHVGELNRLTD 179
 DB 118 YSGIGAGVLAAYIQVSWFCLAGRQILKIRKQFFHAIMRQEIWFVDVHVGELNRLTD 177
 QY 180 DVSKINGIGDKIMFTQSTATFTTGTGIVGTGRGWKLTIVILALSPVLGLSAAIWAHLS 239
 DB 178 DVSKINOVIGDKIMFTQSTATFTTGTGIVGTGRGWKLTIVILALSPVLGLSAAIWAHLS 237
 QY 240 SFTDKELLAYAKAGAVAEVLAARTVAFGGQKKELERNKNLEAKGIGKAITANI 299
 DB 238 SFTDKELLAYAKAGAVAEVLAARTVAFGGQKKELERNKNLEAKGIGKAITANI 297
 QY 300 SIGAAFLIYASALAFWYGTSLVSSEYSIGQVLTIVFFSVLIGAFSIGQASPSIEAFAN 359
 DB 298 SIGAAFLIYASALAFWYGTSLVSSEYSIGQVLTIVFFSVLIGAFSIGQASPSIEAFAN 357
 QY 360 ARGAAEYEFKIIDNKPSIDSYKSGHKPDNIKGNLEFKNVHFSYPSRKEVKILKGLNKV 419
 DB 358 ARGAAEYEFKIIDNKPSIDSYKSGHKPDNIKGNLEFKNVHFSYPSRKEVKILKGLNKV 417
 QY 420 QSGQTVALVGNSSGCKSTTVQLMQRLYDPTDMVCIDQDITRINVRHLREITGVWSQEP 479
 DB 418 QSGQTVALVGNSSGCKSTTVQLMQRLYDPTDMVCIDQDITRINVRHLREITGVWSQEP 477
 QY 480 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPKFTLTVGERGAQSGGQK 539
 DB 478 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPKFTLTVGERGAQSGGQK 537
 QY 540 RIATARALVRNPKIILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 599
 DB 538 RIATARALVRNPKIILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 597
 QY 600 IAGFDDGVIVEKGNHDELMKEGIYFKLVTMTQRTNTELENAATGESKSDALEMSPKD 659
 DB 598 IAGFDDGVIVEKGNHDELMKEGIYFKLVTMTQRTNTELENAATGESKSDALEMSPKD 657
 QY 660 SGSSLLKRSTRSRSHAPQODRKLGTREDNLNENPPVSWFRILKLNSTWPFVVGIFC 719
 DB 658 SRSSLLKRSTRSRVSGSOADRKLSTKEALDESIPPSWFRIMKLNLTWPFVVGVC 717
 QY 720 AINGLQPAFISIFRSIGITFREDPETRQRNSMFSVLVLGUISFTFFLQGTFF 779
 DB 718 AINGLQPAFAIFISKIIGVTRDDPETRQRNSLFLALGUISFTFFLQGTFF 777
 QY 780 GRAGEILRLRYMYFRSMRLQDVSWFDDPKNTTALTRLANDAAQVKGAGSLAVIT 839
 DB 778 GRAGEILTRLRYMYFRSMRLQDVSWFDDPKNTTALTRLANDAAQVKGAGSLAVIT 837
 QY 840 QNANLGTGIIISLYGWQLTLLLAIVPIIAIAGVWEMKMLSGQALKDKELEGAGKIA 899
 DB 838 QNANLGTGIIISLYGWQLTLLLAIVPIIAIAGVWEMKMLSGQALKDKELEGAGKIA 897
 QY 900 TEAIENFRVVSLSLTREQPEYMYAQSLOVPYRNSLRKAHIFGVSPSTQAMMYFYAGCF 959
 DB 898 TEAIENFRVVSLSLTREQPEYMYAQSLOVPYRNSLRKAHIFGVSPSTQAMMYFYAGCF 957
 QY 960 RGCAYLVANEFNFQDVLVLSAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMITEKSPL 1019
 DB 958 RGCAYLVANHLMSFEDVLVLSAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMITEKPL 1017
 QY 1020 IDSYSPLGLKPNTEGNTVFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS 1079
 DB 1018 IDSYSPEGLMPNTEGNTVFGEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS 1077
 QY 1080 TVVQLLERYDPLAGSVLDGKEIKHLNVQWIRAHLGIVSQBPILFDGCSIAENIAYGNS 1139
 DB 1078 TVVQLLERYDPLAGSVLDGKEIKHLNVQWIRAHLGIVSQBPILFDGCSIAENIAYGNS 1137
 QY 1140 RVVSHEEIMOAAKEANIHFIEITLPEKYNTVRGDKGTQSLGGQKORIAARALVRQPHIL 1199
 DB 1138 RVVSQEEIVRAAKEANIHFIESLPNKYSTKVGDKGTQSLGGQKORIAARALVRQPHIL 1197
 QY 1200 LLDATSAIDTESEKVVQALDKAREGRTCVIAHRLSTIQNADLIVVFQNGKVKHEGTH 1259
 DB 1198 LLDATSAIDTESEKVVQALDKAREGRTCVIAHRLSTIQNADLIVVFQNGKVKHEGTH 1257
 QY 1260 QLLAOKGIYFSMVSVQAGAKR 1281
 DB 1258 QLLAOKGIYFSMVSVQAGTKR 1279

Db 718 AINGLOPAFAIIFSKIIGVETRIDDDPETKQNSNLSFLFLAIGIISFITFFLOQFTF 777
Qy 780 GKAGEILTKRLRYMVFERSMLRODYSWFDDPKNTTGALTTRLANDAAQVKGAIGSRLAVIT 839
Db 778 GKAGEILTKRLRYMVFERSMLRODYSWFDDPKNTTGALTTRLANDAAQVKGAIGSRLAVIT 837
Qy 840 QNIANLGTGIIISLIYQWQLTLLIAIIVPIIAIAGVVMKMLSGOALKDKKKELEGAGKIA 899
Db 838 QNIANLGTGIIISFIYQWQLTLLIAIIVPIIAIAGVVMKMLSGOALKDKKKELEGAGKIA 897
Qy 900 TEATENFTVVSRLREQFEYMYAQSLQVYPYRNSLRKAHIFGVSFSTQAMMYFSYACGF 959
Db 898 TEATENFTVVSRLREQFEYMYAQSLQVYPYRNSLRKAHIFGVSFSTQAMMYFSYACGF 957
Qy 960 RFGAYLVANEFMNFQDVLVFSIAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPL 1019
Db 958 RFGAYLVAKLMSFEDVLVFSIAIVFGAMAVGVSSFPADYAKAKISAAHIMIIEKTPL 1017
Qy 1020 IDSYSPHGLKPNLTLEGNVTFNEVVNFYPTREDIPVLOGLSLEVKKGOTLALVSSGCGKS 1079
Db 1018 IDSYSTGLMNPNTLEGNVTFGEVNFYPTREDIPVLOGLSLEVKKGOTLALVSSGCGKS 1077
Qy 1080 TVQOLLERYDPLAGSLIDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSTAENIAYGDNS 1139
Db 1078 TVQOLLERYDPLAGSLIDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSTAENIAYGDNS 1137
Qy 1140 RVVSHEETMOAAKEANLHHFTETLPEKYNTFVGDKGTQLSCGQKORIAIARALVROPHIL 1199
Db 1138 RVVSQEEIVRAAKEANLHAFIESLPNKYSTKVGDKGTQLSCGQKORIAIARALVROPHIL 1197
Qy 1200 LLDATSALDTESEKVVQVEALDKAREGTCIVIAHRLSTIQNADLI VVFONGVKVKEGTH 1259
Db 1198 LLDATSALDTESEKVVQVEALDKAREGTCIVIAHRLSTIQNADLI VVFONGVKVKEGTH 1257
Qy 1260 QOLLAQKGIYFSMVSVQAGAKR 1281
Db 1258 QOLLAQKGIYFSMVSVQAGTKR 1279

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:45:15 ; Search time 3803.81 Seconds
(without alignments)
7047.385 Million cell updates/sec

Title: US-09-672-725C-25

Perfect score: 6473

Sequence: 1 MDPEGGRKGSARKFNWKGK.....LLAQKGIYFSMSVQAGAKR 1281

Scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09672725/runat_04112002_124341_17977/app_query.fasta_1.6435
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725.cgn_1.11409_@runat_04112002_124341_17977 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	6473	100.0	4279	6	AX105080 Sequence
2	6468	99.9	4279	6	AX105078 Sequence
3	6466	99.9	4279	6	AX105082 Sequence
4	6463	99.8	4279	6	AX105057 Sequence
5	6454	99.7	4045	12	AF269224 Synthetic
6	6425.5	99.3	4317	4	AF045016 Canis fam
7	6425.5	99.3	4317	4	AX105059 Sequence
8	6410.5	99.0	3934	4	AF419568 Canis fam
9	5866	90.6	3860	6	AX322787 Sequence
10	5863	90.6	4378	6	E02326 Multidrug r
11	5862	90.6	4186	6	AX108654 Sequence
12	5859.5	90.5	4195	6	AX108656 Sequence
13	5857	90.5	3860	6	AX322789 Sequence
14	5857	90.5	4646	6	AX336420 Sequence
15	5857	90.5	4646	6	AX336708 Sequence
16	5857	90.5	4646	6	I49610 Sequence 2
17	5857	90.5	4646	9	M14758 Homo sapien
18	5857	90.5	4669	6	I08557 Sequence 3
19	5857	90.5	6505	6	AR028671 Sequence
20	5857	90.5	8630	6	AX012320 Sequence
21	5857	90.5	8630	6	AX012321 Sequence
22	5857	90.5	9318	6	AR028672 Sequence
23	5849.5	90.4	3988	6	AX024454 Sequence
24	5849.5	90.4	4192	9	AF016535 Homo sapi
25	5830	90.1	4669	6	AR091275 Sequence
26	5815.5	89.8	4264	6	AR051647 Sequence
27	5815.5	89.8	4264	6	AR051650 Sequence
28	5797	89.6	4669	6	AR055785 Sequence
29	5685.5	87.8	3858	4	OAU78609 Ovis aries
30	5667.5	87.6	4296	10	CRUPGPI
31	5656.5	87.4	4304	10	CRUPGPI1165
32	5641	87.1	4788	6	AX322793 Sequence
33	5641	87.1	4924	10	MUSMDR1A
34	5638	87.1	4356	10	MUSMDR1A
35	5634	87.0	4927	10	AF257746
36	5612	86.7	4323	10	AF286167
37	5521.5	85.3	3489	4	AB029153
38	5395.5	83.4	3987	10	CRUPGPI1185
39	5301	81.9	4189	6	AX322791 Sequence
40	5301	81.9	4298	10	MUSMDR
41	5291.5	81.7	4233	6	AR123273 Sequence
42	5283	81.6	4280	10	CRUPGPII
43	5231.5	80.8	4254	10	RATMDRM
44	4995	77.2	3682	10	MUSPGLY1A
45	4915.5	75.9	3924	6	AX024455 Sequence

ALIGNMENTS

RESULT 1	AX105080	AX105080	4279 bp	DNA	linear	PAT 30-APR-2001
LOCUS	Sequence	24	from Patent WO0123540.			
DEFINITION	Sequence	24	from Patent WO0123540.			
ACCESSION	AX105080					
VERSION	AX105080.1	GI:13921230				
KEYWORDS	dog.					
SOURCE	Canis familiaris					
ORGANISM	Canis familiaris					

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 4279)
REFERENCE AUTHORS Stocker,P.J., Steimel-Crespi,D.T., Crespi,C.L., Reif,T.C. and Patten,C.J.
P-glycoproteins and uses thereof
TITLE Patent: WO 0123540-A 24 05-APR-2001;
JOURNAL GENTEST CORPORATION (US)
FEATURES
source Location/Qualifiers
1..4279
/organism="Canis familiaris"
/db_xref="taxon:9615"
17..3862
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37722.1"
/db_xref="GI:13921231"
/translation="MDPEGGKGRSAEKNFWMKSKKKKKKKKKTSTFTAFMRYSN
WLDRLVLMVGTWAAIIGHGAALPLMLVFNMTDSFANAGISRNKTFPVIINESITNNT
OHPINHLPEEMTYAYVYSGIGAGVLVAAIQVSPWCLAAGROILKIRKOPFHAIIMRO
ETGWDHVDVGLNRLPDDVSKINEGDKIGKEFOSIAETFEFGFVFRGKWLTL
VLAISPVLGSAALAKILLSFTDKELLATAKAGAAVEVLAALRTVIAEGGKKEL
ERYNNLEAGKIGIKKAITANISIGAAPLLIYASYALAFYATWGTSLVLSSEYSQVVL
TVFVSLVIGAFSIGQSPSIEAFANAGAAVEIFKIIDNKPISIDSYSKSGHKPNIGK
NLEFNHVFSPSRKEVKILGNLKVQSGQTVLVGNSGGCKSTVQLMQRLDYPTD
GMVCIDGDIRINVRHLREITGVVSPVLPFATTIAENIRYGRENVMTDETERKAVE
ANAYDFIMLPNKFTLVGERGAQLSGGQKORIAARALVRNPKILLDEATSLDTE
SEAVQVALDKARKGRTTIVIAHRLSTVNRADVIAGFDGVIYVEKGNHDELMKEGLY
FKLVTMTNGNEIELENATGESKSDALEMSPKDSGSLIKRRSTRSIIHAPQGDQR
KLGTEDNLNTEPVPYFWRLKLNSTEMPVYFVFCALINGLOPAPSIIFSRILGI
FRDEPDETQRKNSNMVFLVLGIIISFTIFLQGFTRKAGELLTKRLRYMFRSM
LRQVSDFFDPKNTGALTRLANDAAOVKAIGSRGLAVITONIANLGTGIIISLIYG
WOLTLALLAIVPIIATAGVEMKLSGOALKDKKELEGAKIATEAIEFNRTVNSLTR
EKQFEWYASLOVYPRNRSKHAHFYGSFSTQAMMYFSVAGCFRFGAYVANEEMN
FODLLVFAIVGAMVQVSSFPDYAKAVSAAHVIMIEKSPLDISTSPHGLKP
NTEGNTVNFVFNYPTRPDIPVQLGSLVLEKQTLALVSGSGCKSTVVDLLERF
YDPLAGSVLIDGKEIKHLNWLRAHLGIVSQEPILFDCSAENATGDNRSVSHSE
IMAAKEANIHFIETLPKYNTRVGDQKQLSGGQKORIAARALVRNPKILLDEA
TSALDTEPESKVVQEALDKAREGRTCTIVIAHRLSTIQNALDIWVFQNGVKKEHGTHQL
LQKGIYFSWVSQAGAKR"

BASE COUNT 1296 a 833 c 1008 g 1142 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6473.00 Matches: 1281
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-672-725c-25 (1-1281) x AX105080 (1-4279)

Qy 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPhetPrLysMetGlyLys 20
Db 17 ATGGATCTCTGAAGGAGCCGTAAGGGAGTGCAGAGAAAGAACTTCTGGAAAATGGGCAA 76
Qy 21 LysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
Db 77 AAAAGTAAAAAAGAGAGAAAGAAAGAAACCAACTGTCAGACGTTTGGCAATGTTT 136
Qy 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLallelle 60
Db 137 CGCTATTCAAATGGCTTGTAGTTCTATATGTTGGTGGGCAATGGCTGCCATCATC 196
Qy 61 HisGlyAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
Db 197 CATGGAGCTGCACCTCCCTCATGATGCTGCTGTTTGGAAACATGACAGATAGCTTTGCA 256
Qy 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 257 AATCGAGGAATTTCAAGAAACAAACATTTTCCAGTTATATAATGAAGAGTATTACGAAC 316

Qy 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120
Db 317 AATACACAAACATTTCATCAACCATCTGGAGAGAAATGACCACGTATGCTATTATTAC 376
Qy 121 SerGlyIleGlyAlaGlyValLeuValAlaIleTyrIleGlnValSerPheTrpCysLeu 140
Db 377 AGTGGGATCGGTGCTGGCGTCTGCTTACATCCAGGTTTCATTCTGGTGCTG 436
Qy 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
Db 437 GCAGCAGGAGACAGATACTCAAAATAGAAAACAAATTTTTCATGCTATCATCGCAG 496
Qy 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
Db 497 GAGATTGCTGGTTGAGTGCATGAGCTTGGGAGCTTAACACCGGCTCACAGACGAT 556
Qy 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
Db 557 GTCTCCAAAATCAATGAAGGAATTTGGCGACAAAATTTGAATGTTCTTCAATCAATAGCA 616
Qy 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
Db 617 ACATTTTTCACCGGTTTATAGTGGGTTTACAGTGGTTCGGAAGCTAACCCCTTGAT 676
Qy 221 LeuAlaIleSerProValLeuGlyLeuSerAlaIleIleTrpAlaLysIleLeuSerSer 240
Db 677 TTGGCCATCAGCCCTGTTCTTGGACTTTTCAGCCGCCATCTGGGCAAGATATACTATCA 736
Qy 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
Db 737 TTTACTGATAAAGAACTCTTGGCTTATGCAAAAGCTGGAGCAGTAGCTGGAAGAAGTCTTA 796
Qy 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
Db 797 GCAGCAATCAGAACTGCTATTGCTTTCGAGCACAAGAAAGAACTTGAAGAGTACAAC 856
Qy 281 LysAsnLeuGluGluAlaLysGlyIleGlyLysAlaIleThrAlaAsnIleSer 300
Db 857 AAAAAATTTAGAAGAGCTTAAAGGAATTTGGGATAAAGAAAGCTATCACGCCCAACATTTCT 916
Qy 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrTyrGlyThr 320
Db 917 ATGTGTCGCGCTTCTTATTGATCATCATCATATGCTCTGGCTTCTGGTATGGGACC 976
Qy 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
Db 977 TCCTTGGTCTCTCCAGTGAATATCTATTGACAAAGTACTCACTGCTCTCTTCTTCTGTA 1036
Qy 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db 1037 TTAATTTGGGGCTTTTAGTATTGGACAGGCATCCCCCAAGCATTTGAAGCAATTTGCAAC 1096
Qy 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
Db 1097 AGAGAGCAGCTTATGAATCTTCAAGATAATTGACATAAACCAGCATTTGACAGCTAT 1156
Qy 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1157 TCGAAGAGTGGACATAAACCAGATAATATTAAGGGAATTTGGAATTTCAAAATGTTTCA 1216
Qy 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1217 TTTCACTTACCTTCTCGAAAAGAGTTAAGATCTTAAAGGGGTCTCAACCTGAAGGTTTCA 1276
Qy 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1277 AGTGGCAGACAGTGGCGCTGGTGGGAACAGTGGCTCGGGAAGAGCAGCCGCTGCAG 1336
Qy 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyClnAspIle 460
Db 1337 CTGATGCAGAGGCTCTATGACCCACAGATGGCATGGTCTGTATTGATGAGCAGACAT 1396
Qy 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnProVal 480

Db	1397	AGGACCAATAATGTAAGGCATCTTCGGGAAATTTACTGGTGGGTGAGTCAGAGCGCTGTG	1456
Qy	481	LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	500
Db	1457	TTGTTTCCACCACGATAGCTGAACACATTCGCTATGGCCGGGAAATGTCCACCATGGAT	1516
Qy	501	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	520
Db	1517	GAGATTGAGAAAGCTGTTAAGGAAGCAATGCTATGATTTTATCATGAAACTACCTAAT	1576
Qy	521	LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg	540
Db	1577	AAATTTGACACTCTGGTGTGGAGAGAGGGGCCACGCTAGTGGTGGACAGAAACACAGA	1636
Qy	541	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	560
Db	1637	ATCGCCATTCCTCGGCCCTGGTTGCGACACCCACAGATTCCTCTGCTGGATGAGCGAACG	1696
Qy	561	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	580
Db	1697	TCAGCTCTGGACACTGAAAGTAGGAGCGGGTTCAGGTGGCCCTGGATAAGGCCAGAAAA	1756
Qy	581	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	600
Db	1757	GGCCGGACTACCAATGTGATAGCTATGCTGTTTGTCTACAGTTCGTAATGCCGATGTCA	1816
Qy	601	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	620
Db	1817	GCTGGYTTTGATGTGGAGTCATGTGGAGAAGGAATCATGATGAACCTCATGAAAGAG	1876
Qy	621	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	640
Db	1877	AAGGCATTTTACTTCAAACTGTGCAAATGCAGACACAGAGGAATGAAATTTGAGTTAGAA	1936
Qy	641	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	660
Db	1937	AAATGCCACTGGTGAATCCAAAAAGTGAAGATGATGCTTTGGAAATGTCTCCAAAAAGATCA	1996
Qy	661	GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln	680
Db	1997	GGGTCCAGTTTATAAAGAAGATCACTCCGAGAGTATACATGCACCAAGGCCAA	2056
Qy	681	AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPheTyr	700
Db	2057	GACAGAAAGCTTGATCAAAAGAGGACTTGAATGAGATGTACCTCCAGTTTCCCTCTCGG	2116
Qy	701	ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla	720
Db	2117	AGGATTTCTGAAGCTGAACCTCAACTGAATGGCCTATTTTGTGGTGGTATATTTTGTGCT	2176
Qy	721	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	740
Db	2177	ATTATTAACGGAGGCTCGCAACAGCATTTTCAATAATATTTTCAAGATTTATAGGATC	2236
Qy	741	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	760
Db	2237	TTTACCCGAGATGAGGATCCTGAAACAAACACACAGATAGTACATGTTTCTGTATTG	2296
Qy	761	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	780
Db	2297	TTTCTAGTCTCTGGAAATATTTCTTTTATTTACATTTTTTCTCCAGGGCTTCACATTTGGC	2356
Qy	781	LysAlaGlyGlnIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	800
Db	2357	AAAGCTGGGGAGATCTCTACTAAGCGGCTTGATACATGTTTTCAGATCCATGCTGAGA	2416
Qy	801	GlnAspValSerTyrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	820
Db	2417	CAGGATGTCAGCTGGTTGATGACCTTAAACACCACTGGAGCATTTGACACACCGGCTT	2476
Qy	821	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	840

Db	2477	GCCAAATGATGGCGCTCAAGTTAAAGGGGGCTATAGGTTCCAGCGCTTGCTGTCAATACCAG	2536
Qy	841	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleYrGlyTrpGlnLeuThr	860
Db	2537	AAATAGCAATCTTGGACAGCATATATATATCTTAATCTATGGTTGGCAATTAAACA	2596
Qy	861	LeuSerGlyGlnAlaLeuValProIleIleAlaIleAlaGlyValValGluMetLysMet	880
Db	2597	CTTTTACTCTTGTAGCAATGTGTACCCATCATGTGCAATAGCAGGAGTTGTTGAAATGAAATG	2656
Qy	881	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr	900
Db	2657	TGTCTGGACAGCACCTGAAAGATAGAAGAGCTAGAGGAGCTGGGAAGATTGCTACA	2716
Qy	901	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluYr	920
Db	2717	GAAGCCATCGAAAACCTTCCGAACCTGTGTCTTTGACTCGGACGACAGAAGTTTGAATAC	2776
Qy	921	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	940
Db	2777	ATGTATGCACAGAGTTTGGCAAGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTTC	2836
Qy	941	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	960
Db	2837	GGGGTCTCATTTCTATCAOCCAGGCAATGATGATATTTTCTATGCTGGCTGTTCOCGG	2896
Qy	961	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe	980
Db	2897	TTTGGTGCTACTTGGTGGCAATGAGTTCATGAACCTTCAGGATGTCTTTTGGATATC	2956
Qy	981	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	1000
Db	2957	TCAGCTATGTCTTGGTGCCATGGCAGTGGGAGGTCAGTTCATTTGCTCCTGACTAT	3016
Qy	1001	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1020
Db	3017	GCCAAAGCCAAAGTATCAGCAGCCACGTCATCATGATCATGTAATAAAAGCCCTCTGATT	3076
Qy	1021	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1040
Db	3077	GACAGCTACAGCCCTCAGCGCCTCAAGCCAAATACGTTGGAAAGAAATGTGACATTTAAT	3136
Qy	1041	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1060
Db	3137	GAGTGTGTGTCAACTATCCACTCGACACAGACATCCCGCTGCTCCAGGGGCTGAGGCCTC	3196
Qy	1061	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1080
Db	3197	GAGGTGAAGAAGGGCCAGACGCTGGCCCTCGTAGTAGCAGTGCCTGTGGGAAGAGACACA	3256
Qy	1081	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1100
Db	3257	GTGTTCAGCTCCTTAGAGCGCTTCTATGACCCCTTGGCTGGTTGAGTGTAAATGTATGGC	3316
Qy	1101	LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1120
Db	3317	AAAGAGATAAAGCACCTGAATCTCCAGTGGCTCCGAGCACACCTTGGCATCGTCTCAG	3376
Qy	1121	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1140
Db	3377	GAGCCCATCTCTTTGACTGCAGCATTCGCGAGAACATTGCTATGAGACACACAGCCGG	3436
Qy	1141	ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle	1160
Db	3437	GTCGTATCATGAAGAGATTATGAGGCGAGCCAGGAGGCCAATATACACCACTTTCATC	3496
Qy	1161	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1180
Db	3497	GAGACTCCCTCAGAAATACAAACCCAGTAGGAGACAAAGAACCCAGCTCTCTGGT	3556
Qy	1181	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1200
Db	3557	GGCCAGAAACAGCGCATTCGCATAGCTCGCGCTCTTGTATAGACGCTCATATTTGGCTT	3616

QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
|||||
Db 3617 TTGCATGAAGCTACATCACTCTGATACAGAAAGTGAAGAGTTGTCCAAAGACCCCTG 3676
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
|||||
Db 3677 GACAAAGCCAGAGAGGCGCACCTGCTATGTGATCGCCACCGCTGTGCCACCATCCAG 3736
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
|||||
Db 3737 AATGCGAGATTAAATAGTGTGTTTCAGAAATGGCAAGTCAAGAGACATGCGCACATCAA 3796
QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
|||||
Db 3797 CAGCTGCTGCCAGAAAGGCATCATTTTTCATGGTCAGTGTCCAGGCTGGAGCAAAG 3856
QY 1281 Arg 1281
|||
Db 3857 CGC 3859
RESULT 2
AX105078
LOCUS AX105078 4279 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 22 from Patent WO0123540.
ACCESSION AX105078
VERSION AX105078.1 GI:13921228
KEYWORDS
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 4279)
AUTHORS Stocker,P.J., Steimel-Crespi,D.T., Crespi,C.L., Reif,T.C. and
Patten,C.J.
TITLE P-glycoproteins and uses thereof
JOURNAL Patent: WO 0123540-A 22 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
source
1. .4279
/organism="Canis familiaris"
/db_xref="taxon:9615"
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37721.1"
/db_xref="GI:13921229"
/translation="WDPEGKRGSAEKNFWKMKKKKNEKKKPTVSTFAMPRYSN
WDLRLMLVGTWAAIIGHAALPLMLVFCNMTDSPANAGISRKTFPVLINESINTNT
QFPHNLEEEMTYAYISGIGAGVLVAAYIQVSEWCLAAGRQILIKRQFFHMRQ
EIGWFDVHDVGLNLTLDVSKINEGIDKIGMFQSIATFTFGTIVFGFRGKMLTL
VILAI5PVLGSLAAITWAKLSTFDKELLAYAGAVAEVLAIRTVIAFGQKKEL
ERYKNLEAKGIGIKKAITANISIGAAFLLIYALAFWYGLSLVSSYSIGQVL
TVF5VLGAF5SISGQASPIEAFANARNAAYEIKIIONKPSIDSYSKSHKPNKIG
NLEFNHVSFSRKEVKILKGLNKLKVSQGTVALVGNSSGCKSTTVOLMQRLDPTD
GWGIDGQDITINVRHUREITGVVSQEPVLFATTIAENIRYGRNVNDEIERAKVE
ANAYDFIMKLPNKFTLVGERGAQLSGGQKRIATARALVNRKILLIDDEAT5ALDTE
SEAVQVADLARKQRTTIVIAHRLSTVRNADYIAGFDGQIVIEKGNHDELMKEKGIY
FKLVTMDNENYPIELNATGSKSESDALEMSPKDSGLKRRSTRSIRHAPQODR
KLGTEDLNENYPIELNATGSKSESDALEMSPKDSGLKRRSTRSIRHAPQODR
FRDEDEPKRONSNNF5VLVLGI5EIIFFLQGTGKAGELLTKRLRYMFRSM
LRQDV5WFDPKNTGALTIRUANDAQQVKGAIGSRNAVITONIANLGTGLIISLYG
WQTLTLLAIVPIAIAGVVEMKMSGLQALDKKELEGAKIATEINFRVTVSLTR
EQKF5YMAQ5LQVYPRN5LRKAHIFGVSF5ITQAMMYFSYAGCFRFGAYLVANEFM
FODVLIVFAIVFGAMAGOVSS5FADYAKAVSAAHVIMIEK5PLD5SY5PHGLKP
NPLEGNVFN5VNNY5P5PDIPVLQGL5LEV5KQGTALV5GSS5CG5LTV5QLLERF
YPLAG5VILDK5EIKHLNVQMLRAHGLTV5Q5EPL5FDCA5AENAY5GDSNRV5V5HEE
I5QAAKEAN5HIF5ETL5PEKYN5RV5DKGT5QL5GQK5ORIA5RALV5RQ5PHILL5DEA
T5ALDTE5SEK5VVO5EAL5DK5GRT5CIV5IAHRL5STIQ5NAL5IV5FONG5VKE5GHT5HQL
LAQKGI5FSM5V5Q5AGAKR"
BASE COUNT 1295 a 833 c 1008 g 1143 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6468.00 Matches: 1280
Percent Similarity: 99.92% Conservative: 0
Best Local Similarity: 99.92% Mismatches: 1
Query Match: 99.92% Indels: 0
DB: 6 Gaps: 0
US-09-672-725C-25 (1-1281) x AX105078 (1-4279)
QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
|||||
Db 17 ATGTATCTCTGAAGAGGCGCTAAGGGAGTGCAGAGAGAACTTCTGGAAAAATGGGCAA 76
QY 21 LysSerLysLysGluLysGluLysLysProThrValSerThrPheAlaMetPhe 40
|||||
Db 77 AAAAGTAAAAAANTGAGAGAAAGAAAGAAACCAACTGTGCAGCAGCTTTGCAATGTT 136
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 60
|||||
Db 137 CGCTATTCAAATTTGGCTTGATAGGTTGTATATGTTGGTGGGACAATGGCTGCCATCATC 196
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
|||||
Db 197 CATGGAGCTGCACCTCCCTCTCATGATGCTGGTGTGGAAACATGACAGATAGCTTTGCA 256
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
|||||
Db 257 AATGCAGGAATTTCAAGAAACAAAACTTTCCAGTTATATAATTAATGAAGTATTACGAAC 316
QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 120
|||||
Db 317 AATACACAACTTTTCATCAACCATCTGGAGAGGAAATGACCACGTATGCTTATATTAC 376
QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
|||||
Db 377 AGTGGATCGGTGCTGGCGTCTGCTGGTGGCTTACATCCAGGTTTCATTCTGGTGCTG 436
QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
|||||
Db 437 GCAGCAGGAAGACAGATACTCAAAATTAGAAAAACAATTTTTTCATGCTATCATCGCAG 496
QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
|||||
Db 497 GAGATTGGCTGGTTGACGTGATGACGTTGGGAGCTTAAACCCGCTCACAGACAT 556
QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
|||||
Db 557 GTCCTCAAAATCAATGAAGGAATTTGGCGACAAAATTTGGAATGTTCTTCAATCAATAGCA 616
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
|||||
Db 617 ACATTTTTCACCGGTTTTATAGTGGGTTTACAGTGGTTGGAAGCTAACCCCTTGTGATT 676
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
|||||
Db 677 TTGGCCATCAGCCCTGTTCTTGGACTTTTCAGCCGCCATCTGGGCAAGAGATACATCTTCA 736
QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
|||||
Db 737 TTTACTGATAAAGAACTCTTGGCTATGCCAATGCGAGCAGTAGCTGGAAGAAGTCTTA 796
QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
|||||
Db 797 GCAGCAATCAGAACTGTGATTGCCCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAAC 856
QY 281 LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
|||||
Db 857 AAAAATTTAGAAGAACCTPAAAGGAATTGGGATAAAGAAAGCTATCACGGCCAACTTCT 916
QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
|||||
Db 917 ATTGGTCCCGCTTTCTTATTGATCTATCATCATATGCTCTGGCTTTCTGTGTGGGACC 976

QY	321	SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal	340
DB	977	TCCTTGGTCTCTCCAGTGAATATTCTATTGGCAAGTACTCAGTGTCTCTTTCTCTGA	1036
QY	341	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	360
DB	1037	TTAATTGGGCTTTAGTATTGCAGAGGCATCCCAAGCATTTGAAGCATTTGCCAAAGCGCA	1096
QY	361	ArgGlyAlaIaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	380
DB	1097	AGAGGACAGCTTATGAATCTTCAAGATAATTGACAATAAACCAAGCATTTGACAGCTAT	1156
QY	381	SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	400
DB	1157	TCGAAGAGTGGACATAAACACAGATAATATTAAAGGAAATTTGGAAATTCAAAATTTGTCAC	1216
QY	401	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	420
DB	1217	TTCAATTACCTTCTCGAAAGAAGTTAAGATCTTAAAGGCTCTCAACTGAAAGTTTCAG	1276
QY	421	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln	440
DB	1277	AGTGGGAGACAGTGGCGCTGGTTGGAAACAGTGGCTCGGGAGAGCAGCACCGTGCAG	1336
QY	441	LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	460
DB	1337	CTGATGCAGAGGCTCTATGACCCACAGATGCATGGTCTGTATTGATGGACAGACATT	1396
QY	461	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	480
DB	1397	AGGACCAATAATTAAGGCATCTTCGGGAAATTACTGGTGGTGAGTCAGGACCGTGTG	1456
QY	481	LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	500
DB	1457	TTGTTTCCACACAGTAGCTAAAAACATTCGCTATGGCCGCGAAAAATGTCAACATGGAT	1516
QY	501	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	520
DB	1517	GAGATTGAGAAACGCTTAAGAGGCCAATGCGCTATGATTTATCATGAACATACCTTAAT	1576
QY	521	LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg	540
DB	1577	AAATTTGCACACTCTGTTGGAGAGAGGGGCCAGCTGAGTGGTGAGCAAGAACAGAGA	1636
QY	541	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	560
DB	1637	ATCGCCATTGCTCGGCCCTGGTTGCGAACCCCAAGATTCTTCGTGGATGAGGCAACG	1696
QY	561	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	580
DB	1697	TCAGCTCTGGACACTGAAGAGTGAAGCTGGTTCAAGTGGCCCTGGATATAGGCCAGAAA	1756
QY	581	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	600
DB	1757	GGCGGCACTACCATTTGATAGTCACTCATGTTGTCTACAGTTCTGTAATGCCGATGTCAAT	1816
QY	601	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	620
DB	1817	GCTGGTTTGAATGATGGATCAATTGGAGAAAGAAATCATGATGAACATCATGAAGAG	1876
QY	621	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	640
DB	1877	AAGGGCATTTACTTCAAACTTGTCAATGCAGACACAGAGAAATGAATTTGAGTTAGAA	1936
QY	641	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	660
DB	1937	AATGCCCACTGGTGAATCCAAAAGTGAAGTGTATGCTTGGAAATGTCTCCAAAAGATTCA	1996
QY	661	GlySerSerLeuIleLysArgSerThrArgSerIleHisAlaProGlnGlyGln	680
DB	1997	GGGTCCAGTTTAATAAAAAGAGATCACTCGAGAGTATACATGATGCCACCAAGGCCAA	2056

QY	681	AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	700
Db	2057	GACAAAGCTTGGTACAAAAGAGGACTTCAATAGAAATGTACCTCCAGTTTCCCTTCGG	2116
QY	701	ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	720
Db	2117	AGGATTCGAAAGCGAACTCAACTGAATAGCCCTATTTTGTGGTGGTATATTTTGTGCT	2176
QY	721	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	740
Db	2177	ATTATATAACGAGGCGCTGCAACACGACATTTTCAATATATTTTCAAGGATATAGGGATC	2236
QY	741	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	760
Db	2237	TTTATCCCGAGATGAGGATCCTGAAACAAAACGACAGAAAGTAGTAACATGTTTCTGTATTG	2296
QY	761	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	780
Db	2297	TTTCTAGTCCCTGGAAATATTCTTTATTTACATTTTCTCCTCAGGGCTTCACATTTGGC	2356
QY	781	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	800
Db	2357	AAAGCTGGGAGATCCTCACTAAGGCGTTCGATACATGGTTTTCAGATCCATGCTGAGA	2416
QY	801	GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	820
Db	2417	CAGGATGTCACTGGTTGTATGACCCTAAAAACACCACTGGAGCATTCACAACCAAGGCTT	2476
QY	821	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	840
Db	2477	GCCAATGATCGGCTCAAGTTAAAGGGCTATAGTTCACAGCTTGC7GTCAATACCCAG	2536
QY	841	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr	860
Db	2537	AATATAGCAATCTTGGACAGGCATTTATATATCTTAATCTATGTTGGCAATTAACA	2596
QY	861	LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet	880
Db	2597	CTTTTACTCTTAGCAATTGTACCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAAATG	2656
QY	881	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr	900
Db	2657	TTGTCTGGACAAGCACTGAAAGATAAGAAGAGCTAGAAAGAGCTGGGAAGATTGCTACA	2716
QY	901	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr	920
Db	2717	GNAGCCATCGAAAACCTCCGAACTGTGTTCTTTGATCTCGGAGCAGAACTTTGAAATAC	2776
QY	921	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	940
Db	2777	ATGTATGCACAGAGTTTGTCAAGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTTC	2836
QY	941	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	960
Db	2837	GGGGTCTCATTTCTATCAACCCAGGCAATGATGTATTTTCTATGCTGGCTGTTCCTCGG	2896
QY	961	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe	980
Db	2897	TTTGGTGCCTACTTGGTGGCAATAGATTCATGAACATTCAGGATGTTCTTTTGGTATTC	2956
QY	981	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	1000
Db	2957	TCAGCTATTGTCTTTGTGGCATGGCAGTGGGCGAGTCAGTTTCATTTGCTCCTGACTAT	3016
QY	1001	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1020
Db	3017	GCCAAAGCCAAAGTATCAGCAGCCCACTCATCATCATCATTTGAAAAAGCCCTCTGATT	3076
QY	1021	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1040
Db	3077	GACAGCTACAGCCCTCAGGCCCTCAAGCCAAATACGTTGGGAAGGAAATGTGACATTAAT	3136
QY	1041	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1060

|||||
Db 497 GAGATTGGCTGGTTTGACGTGCATGACCTGGGGAGCTTAACACCCGGCTCACAGCCAT 556
Qy 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
Db 557 GTCTCCAAAATCAATGAAGGAATTTGGCGACAAAATTTGGAATGTTCTTTCAATCAATAGCA 616
Qy 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTyrLysLysLeuThrLeuValIle 220
Db 617 ACATTTTTCACCGGTTTATAGTGGGTTTACACGTGGTTGGAAGCTTAACCCCTTGTAAT 676
Qy 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTyrAlaLysIleLeuSerSer 240
Db 677 TTGGCCATCAGCCCTGTTCTTGGACTTTCAGCGCCATCTGGCAAGAGATACTATCTCA 736
Qy 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
Db 737 TTTACTGATAAAGAACTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 796
Qy 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
Db 797 GCAGCAATCAGACTGTGATGGCTTTGGAGGACAAAGAAGAAGAACTTGAAGGTACAAC 856
Qy 281 LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
Db 857 AAAAATTTAGAAGAAGCTTAAAGGAATTTGGGATAAAGAAAGCTATCACGGCCAACTTCT 916
Qy 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrTyrGlyThr 320
Db 917 ATTGGTCGGCTTCTTATTGATCATGATCATATGCTCTGGCTTTCTGGTATGGACC 976
Qy 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
Db 977 TCCCTGGTCTCTCCAGTGAATATCTATTGGACAAGTACACTGCTCTCTTTCTTCGTA 1036
Qy 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db 1037 TTAATTGGGGCTTTTAGTATTGGACAGCATCCCCAAAGCATTGAAGCATTTCGCAACGCA 1096
Qy 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
Db 1097 AGAGGACAGCTTATGAATCTTCAAGATAATTGACAAATTAACCAAGCATTTGACAGCTAT 1156
Qy 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1157 TCGAAGAGTGACATAAACACAGATAATATTAAAGGAATTTGGAATTCAAAATGTTTCA 1216
Qy 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1217 TTCAGTTACCCCTTCTCGAAAAGAAGTTAAGATCTTAAGGGTCTCAACCTGAAGGTTTCA 1276
Qy 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1277 AGTGGGACAGACAGTGGCGCTGGTTGGGAACAGTGGCTTCGGGAAGACAGACGCGTGCAG 1336
Qy 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db 1337 CTGATGACAGGCTCTATGACCCACAGATGGCATGGTCTGTATTGTATGGACAGGACATT 1396
Qy 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db 1397 AGGACCATAAATGAAGGCATCTTCGGGAAATTTACTGGTGGTGTGAGTCAGGACCGTGTG 1456
Qy 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1457 TTGTTTCCACACAGATPAGCTGAAAACATTCGCTATGGCGCGGAAAATGTCAACATGGAT 1516
Qy 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1517 GAGATTGAGAAGCTGTTAAGGAAGCAATGCCCTATGATTTTATCATGAAACTACCTAAT 1576
Qy 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
|||||

Db 1577 AAATTTGACACTCTGGTTGGAGAGAGAGGGCCCGACGTGAGTGGTGACAGAAACAGAGA 1636
Qy 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
Db 1637 ATGCCATTCTCGGGCCCTGGTTGCAACCCCAAGATTCTTCTGCTGATGAGCAACG 1696
Qy 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1697 TCAGCTCGACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCTGGATAGGCCAGAAA 1756
Qy 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1757 GGCGGAGACTACCATTTGTGATAGCTCATCGTTCTTACAGTTCGTAATGCCGATGTCATT 1816
Qy 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1817 GCTGGTTTGTATGATGAGAGTCATTTGGAGAAAAGAAATCATGATGAATCATGAAAGAG 1876
Qy 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1877 AAGGGCATTTTACTTCAAACTTGTCACATGCACAGAGAGAAATGAAATTTGATGTAGAA 1936
Qy 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1937 AATGCCACTGGTGAATCCAAAAGCTGAAAGTGATGCCCTTGGAAATGCTCCAAAAGATTCA 1996
Qy 661 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db 1997 GGGTCCAGTTTAATAAAAGAGATCAACTCGCAGGAGTATACATGCACCAAGGCCAA 2056
Qy 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTyr 700
Db 2057 GACAGAAAGCTTGGTACAAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCTCTCTGG 2116
Qy 701 ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 720
Db 2117 AGGATTCTGAAGCTGAACCTCAACTGAATGGCCTTATTTTGTGGTGTATATTTTGTGCT 2176
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2177 ATTATAACGGAGCGCTGCAACCCAGCATTTTCAATAATATTTTCAAGGATTTATAGGATC 2236
Qy 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2237 TTTTACCCGAGATGAGGATCCTGAAACAAACGACAGAAATAGTAACATGTTTCTGTATG 2296
Qy 761 PheLeuValLeuGlyIleIleSerPheIleThrPheLeuGlnGlyPheThrPheGly 780
Db 2297 TTTCTAGTCTCTGGAATTTATTTCTTTTATACATTTTCTCCAGGGCTTCACATTTGGC 2356
Qy 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAAGCTGGGAGATCTCTCACTAAGCGGCTTCGATACATGGTTTTCAGATCCATGCTGAGA 2416
Qy 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGGATGTGAGTGGTGGTTGATGACCCCTAAAACACCCACTGGAGCATTTGACACACAGGCTT 2476
Qy 821 AlaAsnAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2477 GCCAATGATCGGCTCAAGTTAAAGGGGTATAGGTTCCAGGCTTGTCTCATTAACCCAG 2536
Qy 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
Db 2537 AATATAGCAAAATCTTGGGACAGCATTTATATATCTTAATCTATGTTGGCAATTAACA 2596
Qy 861 LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTCTACTCTTAGCAATTTGATCCCATCATTCGAATAGCAGGAGTGTGTTGAAATGAAATG 2656
Qy 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
Db 2657 TTGCTGGACAGCACTGAAAGATAAAGAAGAGCTAGAAGGAGCTGGAGAGATTGCTACA 2716
|||||

```
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
|||||
Db 2717 GAAGCCCTCGAAACATTCGAACTGTTGTTCTTTGACATCGGACGACAAAGTTTGAATAC 2776

QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
|||||
Db 2777 ATGTATGCACAGATTGCAAGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTTC 2836

QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
|||||
Db 2837 GGGGTCTCATTTCTATCACCGCAATGATGATATTTTCTCATGCTGCTGTTTCCGG 2896

QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
|||||
Db 2897 TTTGGTGCTACTTGGTGCAATGAGTTCATGAACTTTCAGATGTTCTTTGGTATTC 2956

QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
|||||
Db 2957 TCAGCTATTGTTCTTGGTGCCATGGCAGTGGGCAGGTCAAGTTCATTTGCTCTGACTAT 3016

QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
|||||
Db 3017 GCCAAAGCCAAAGTATCAGACGCCACGTCATCATGATCATTTGAAAGACCCCTCTGATT 3076

QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
|||||
Db 3077 GACAGCTACAGCCCTCACGCCCTCAGCCCAATACAGTTTGGAGGAAATGTGACATTTAAT 3136

QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
|||||
Db 3137 GAGGTGCTGTTCAACTATCCCACTCGACACGACATCCCGTCTCCAGGGGCTGAGCCTC 3196

QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
|||||
Db 3197 GAGGTGAAGAAGGCCAGACGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGACACA 3256

QY 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
|||||
Db 3257 GTTGTTCAGCTCTCAGCGCTTCTATGACCCCTTGCTGGCTGTTTCAGTGTCAATTGATGC 3316

QY 1101 LysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSerGln 1120
|||||
Db 3317 AAAGAGATAAAGCACCTGAATGTCCAGTGGCTCCGAGCACACTGGGCATCTGTGCTCAG 3376

QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
|||||
Db 3377 GAGCCCATCTGTTGACTGCAGCATTCGCGAGAACATTCGCTATGGAGACACACGCCG 3436

QY 1141 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1160
|||||
Db 3437 GTCGTATCATGAAGAGATTGTGACGAGCGCAAGCGCAACATACACACCACTTCATC 3496

QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
|||||
Db 3497 GAGACACTCCTCCTGAGAAATACAAACACACAGATGAGGAGACAAAGAACCCAGCTCTCTGT 3556

QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
|||||
Db 3557 GGCAGAAACAGCGCATTTGCCATAGCTCGCGCTCTTGTGTAGACAGCCTCATATTTGCTT 3616

QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
|||||
Db 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAAGAGCCCTG 3676

QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
|||||
Db 3677 GACAAAGCCAGAGAGGCCGCCACTGCTGATGTGATCGCCACCGCTGTGCCACCATCCAG 3736

QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
|||||
Db 3737 AATGCAGATTTAATAGTGGTGTTCAGATGGCAAGTCAAGGAGCATGCGACACATCAA 3796
```

```
QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
|||||
Db 3797 CAGCTGCTGGCCAGAAAGCATCTATTTTCCATGTCAGTGTCCAGGCTGGAGCAAG 3856

QY 1281 Arg 1281
|||
Db 3857 CGC 3859

RESULT 4
LOCUS AX105057 4279 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 1 from Patent WO0123540.
ACCESSION AX105057
VERSION AX105057.1 GI:13921209
KEYWORDS .
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 4279)
AUTHORS Stocker,P.J., Steimel-Crespi,D.T., Crespi,C.L., Reif,T.C. and
Patten,C.J.
TITLE P-glycoproteins and uses thereof
JOURNAL Patent: WO 0123540-A 1 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
source
1. .4279
Location/Qualifiers
/organism="Canis familiaris"
/db_xref="taxon:9615"
17..3862
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC37718.1"
/db_xref="GI:13921210"
/translaton="MDPEGGRKSAEKNFWMGKSKNKKKPKKPTVSTFAMFRYSN
WLDRLVLYGTMAIIGHGALPLMLVFGNMTDFSPANAGISRNKTPPVIINESITNNT
QHPINLHEENTYAYIYSGIGAGLVAAVLIQVSWCLAGBQILAKIRQFHAIMRQ
EIGWDFVDHVGELNRLTDVDSKINGIDKIGKGFHSHSIATPTFTGIVGFTGKWLTL
VILASPVGLSAAIWAKLISSTDELLAYAGAVAEVLAARIVTAFGGKAL
ERYNNLEAGKIGIKKAITANSIGAAFLIYASVALAFWYGTSDLVSEYSIGQVL
TVFSLVIGAFSIGQASPSIEAFANAGAAEYFKIDNKPSIDSYSGSHKPDNKG
NLKFNKVFSPSRKEVKILKLVNKGQSVQTVLVNCGSGKSTVQMLQRLDPTD
GMCIGODIRTINVRHLREITGVSOEPLFATTIENIRYGRNWTMDIEKAVKE
ANAYDFIMKLPNKFEDLVGERGQALSGGQKORIAIARALVRNPKILLDEATSDLTE
SEAVQVALDKARKRTTIVIAHRLSTVIAGFDGVIWEKSHNDELMEKEGIY
FKLVTMTGRNEILENATGESKESDALESPKDSGLIKKRSRTRSIHAFQGDRI
KLGTEDENVPVPSFWRLKLNSTEWPFYVVGIFCAIINGGLQFAPSIIIFSRILGI
FTRDEDEPTKQNSNMFVSLFLVLGIIISFTIFFLQGTGKAGEIILTKRLRYMFRSM
LRDQVSWFDDPKNTTGALTRLANDAAQVKAIGSRILAVITQIANLGTIISLIYG
WOLTLILAIVPIIATAGVVKMLSQALKDKKLEKAGKIATEIENFRVTVSLTR
EQKFYMAQSLOVPYRNSLRKAHIGVSESTQAMMYSYACGFRGAYLVANEFMN
FQDLVLYSAIVFGAMAVGVSSFPADYAKAVSAAHVIMIIEKSPILDSYPHGLKP
NLTEGNVTFNEVFNTPRDIPIVLOGLSLEVKKGQTLALVSGSGCKSTVVOQLERF
YDPLAGSVLIDGKEIKHLNQWLRAHLGIVSOEPIIFDCSIAENIAYGNSRVVSHEE
IMQAAEKANIHETIPEKYNTRVGDKGTQLSGQKORIAIARALVRPQHILLDEA
TSALDTSEKVVQOEALDKAREGRTCTIVIAHRLSTIQNALDILVVFQNGKVEKHTHOOL
LAOKGIYSVMYSVQAGKR*
BASE COUNT 1294 a 834 c 1008 g 1143 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6463.00 Matches: 1279
Percent Similarity: 99.84% Conservative: 0
Best Local Similarity: 99.84% Mismatches: 2
Query Match: 99.85% Indels: 0
DB: 6 Gaps: 0
US-09-672-725C-25 (1-1281) x AX105057 (1-4279)
QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTyrLysMetGlyLys 20
|||||
Db 17 ATGGATCTCTGAAGGAGCGCGTAAGGGGAGTGCAGAGAAAGAACTTCTTGAAATGGGCAA 76
```

QY 21 LysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
 DB 77 AAAAGTAAAAAATGAGAGAAAGAAAGAAACCAACTGTCAGCACGTTTGGCAATGTTT 136
 QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 60
 DB 137 CGCTATTCAAAATGGCTGTGATAGGTTGTATATGTTGGTGGGACAATGGCTGCATCATC 196
 QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
 DB 197 CATGGAGTCGACCTCCCTCATGATGCTGTTTGGAAACATGACAGATAGCTTTGCA 256
 QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
 DB 257 AATCAGAGAAATTCAGAAACAAACCTTTCCAGTTATATTAATGAAGATATTACGAAC 316
 QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 120
 DB 317 AATACACAACATTTTCATCAACCATCTCGAGAGGAAATGACCAGTATGCTATTATTAC 376
 QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
 DB 377 AGTGGGATCGGTGCGCTGCTGCTGCTTACATCCAGGTTTCATTTCTGGTGCCTG 436
 QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
 DB 437 GCAGCAGAAAGACAGATACTCAAAATTAGAAACAAATTTTTCATGCTATCATGCCAGAC 496
 QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
 DB 497 GAGATTGGCTGGTTGACGTGTCATGACGTTGGGAGCTTAACACCGCGCTCACAGACGAT 556
 QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
 DB 557 GTCTCCAAATCAATGAAGAAATGGCGACAAATTTGGAATGTTCTTCACTCAATAGCA 616
 QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
 DB 617 ACATTTTTCACCGGTTTATAGTGGGTTTACACGTGGTGGAGCTAACCCCTTGTGATT 676
 QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
 DB 677 TTGGCCATCAGCCCTGTTCTGGACTTTCAGCCGCCATCTGGGCAAGATACTATCTTCA 736
 QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
 DB 737 TTTACTGATTAAGAAGCTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 796
 QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
 DB 797 GCAGCAATCAGAACTGTGATTGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAAC 856
 QY 281 LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
 DB 857 AAAAAATTTAGAAGAAGCTAAGGAATGGGATAAAGAAAGCTATCACGCCCAACATTTCT 916
 QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
 DB 917 ATTGGTGGCCCTTCTATTGATCTATGCATCATATGCTGCTGGCTTTCGGTATGGGACC 976
 QY 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
 DB 977 TCCTTGGTCTCTCCAGTGAATATTTATTTGGACAAGTACTCACTGCTCTCTTCTTCTGTA 1036
 QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
 DB 1037 TTAATTTGGGCTTTTATGATTGGACAGGATCCCAAGCATTTGAAGCATTTGCAACAGCA 1096
 QY 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
 DB 1097 AGAGGAGCAGCTTATGAAATCTTCAAGATAATTTGACAATAAATCAACCAAGCATTGACAGCTAT 1156

QY 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
 DB 1157 TCGAAGATGGACATTAACACAGATAATATTAAAGGAAATTTGGAATTTCAAAATGTTTAC 1216
 QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
 DB 1217 TTCAGTTACCTTCTCGAAAAGAAAGTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTACAG 1276
 QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
 DB 1277 AGTGGGCACACAGTGGCGCTGTTGGGAACAGTGGCTGCGGGAAGACGACCGCTGCAG 1336
 QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
 DB 1337 CTGATCCAGAGCTCTATGACCCACAGATGCGATGGTCTGTATTGATGCGACAGACATT 1396
 QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
 DB 1397 AGSACCATAAATGTAAGGCATCTTCGGGAAATTTACTGGTGTGGTGGTGGTGGTGGTGG 1456
 QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
 DB 1457 TTGTTTGGCCACCACGATAGCTGAAAACATTCGCTATGGCCGCGAAATGTCACCATGAT 1516
 QY 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
 DB 1517 GAGATTGAAAGCTGTTAAGGAGCCATGCTATGATTTTATCATGAACTACTCTAAT 1576
 QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
 DB 1577 AAATTTGACACTCTGTTGGAGAGAGGGGCCAGCTGAGTGGTGGAGAGAGAGAGAGAG 1636
 QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
 DB 1637 ATCGCATTTGCTCGGGCCCTGTTTCGCAACCCCAAGATTTCTTCTGCTGGATGAGCAACG 1696
 QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
 DB 1697 TCAGCTCTGGACACTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1756
 QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
 DB 1757 GGCCGAGCTACCATTTGTATAGTCTCATCTTGTCTACAGTTTCGTAATCCCGATGTCATT 1816
 QY 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
 DB 1817 GCTGTTTTGATGATGGAGTCAATGTGGAGAAGGAAATCATGATGAATCATGAAAGAG 1876
 QY 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
 DB 1877 AAGGCATTTACTTCAAACTTGTCAATGTCAGACAGAGGAAATGAAATTTGAGTTAGAA 1936
 QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
 DB 1937 AATGCCACTGGTGAATCCAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1996
 QY 661 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
 DB 1997 GGGTCCAGCTTTAAWAAAAAGAGATCAACTCGAGGAGTATACATGCACCAAGGCGAA 2056
 QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
 DB 2057 GACAGAAAGCTGGTACAAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCTCTGG 2116
 QY 701 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
 DB 2117 AGGATTTCTGAAGCTGAACCTCAACTGAATGGCTTATTTTGGTGGTATATTTTGTGCT 2176
 QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
 DB 2177 ATTATAAACGGAGGCTCGAACCCAGCATTTTCAATAATATTTTCAAGGATTTATAGGATC 2236
 QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760

|||||
Db 2237 TTTACCGAGATGAGGATCCTGAACAAAAACGACAGAATAGTAACATGTTTCTGTATTG 2296
Qy 761 PheLeuValLeuGlyIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2297 TTTCTAGTCTTGGAAATTAATTTCTTTTATTACATTTTCTCCACAGGGCTTCACATTTGGC 2356
Qy 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAAGCTGGGGAGATCCTCACTAAGCGGCTTCGATACATGGTTTTCAGATCCATGCTGAGA 2416
Qy 801 GlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGGATGTCAGCTGGTTGATGACCTTAAACACCACTGGAGCATTCGACAAACAGGCTT 2476
Qy 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2477 GCCAATGATGCGGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGGCTCATTAACCCAG 2536
Qy 841 AsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
Db 2537 AATATAGCAAAATCTGGGACAGCATATATATCTTAATCTATGTTGGCAATTAACA 2596
Qy 861 LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTACTCTTAGCAATTTGACCATCATTCGAATAGCAGGAGTTGTTGAAATGAAATG 2656
Qy 881 LeuSerGlyGlnAlaLeuLysAspLysGlyLeuGluGlyAlaGlyLysIleAlaThr 900
Db 2657 TTGCTGGAACAACACTGAAAGATAGAAAGAGCTAGAGGAGCTGGGAAGATTGCTACA 2716
Qy 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2717 GAAGCATCGAAACCTCCGAACCTGTTGTTCTTGACTCGGAGCAGAAGTTTGAATAC 2776
Qy 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2777 ATGTATGCACAGAGTTTGAAGTACCATACAGAAATCTCTTTGAGGAAAGCACACATCTTC 2836
Qy 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2837 GGGGTCTCATTTCTATACCCAGGCAATGATGATTTTCTCATGCTGGCGTTCGG 2896
Qy 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 980
Db 2897 TTTGGTGCCTACTTGTGGCAATGAGTTCATGAACCTTTTCAGSATGTTCTTTTGTATTTC 2956
Qy 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
Db 2957 TCAGCTATTGTCTTTGTGGTGCATGGCAGTGGGCGAGTTCATTTGCTCCTGACTAT 3016
Qy 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
Db 3017 GCCAAGCCCAAGATATCAGACGCCACAGTCATCATGATCATTTGAAAAAGCCCTTGAT 3076
Qy 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3077 GACAGCTACAGCCCTCACGGCTCAAGCCAAATACGTTTGAAGGAAATGTGACATTTAAT 3136
Qy 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3137 GAGGTGCTGTCAACTATCCCACTCGACCAAGACATCCCGTGTCCAGGGGTGAGCCCTC 3196
Qy 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
Db 3197 GAGGTGAAGAGGGCCAGACCGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGACCACA 3256
Qy 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
Db 3257 GTTGTTCAGCTCCTAGACGCTTCTATGACCCCTTGGCTGGTTCAGTGTGCTAATGATGC 3316
Qy 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisGlyIleValSerGln 1120
|||||

Db 3317 AAAGAGATAAAGCACCTGAATGTCCAGTGGCTCCGAGCACACACTGGCATCTGTCTCAG 3376
Qy 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
Db 3377 GAGCCCATCTGTTGACTGCAGCATTCGCCGAGAACATTGCCCTATGGAGACAACAGCCGG 3436
Qy 1141 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1160
Db 3437 GTCGTATCACATGAAGAGATTATGCAGCAGCCCAAGGAGGCCAACATACACCACTTCATC 3496
Qy 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
Db 3497 GAGACATCCCTGAGAATAACACACCGAGTAGGACAAAGAACACCCAGCTCTCTGGT 3556
Qy 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
Db 3557 GGCAGAAACAGCGCATTCGCATAGCTCGCGCTCTTGTAGACAGCTCATATTTTGCCT 3616
Qy 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
Db 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAAGGTTGTCCAAGAGCCCTG 3676
Qy 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
Db 3677 GACAAAGCCAGAGAGCCGCGCACCTGCATGTGTATGCCCCACCGCTGTCCACCATCCAG 3736
Qy 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
Db 3737 AATCAGATTTAATAGTGGTGTTCAGAAATGGCAAGTCAAGGAGCATGGCACACATCAA 3796
Qy 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
Db 3797 CAGTGTGGCCCAAGAGGCATCTATTTTCCATGTGTAGTCCAGGTGTCAGGCTGGAGCAAAG 3856
Qy 1281 Arg 1281
Db 3857 CGC 3859
|||
RESULT 5
AF269224 4045 bp DNA linear SYN 04-JUL-2000
LOCUS Synthetic construct Canis familiaris his-tagged-multidrug
DEFINITION resistance glycoprotein gene, complete cds.
ACCESSION AF269224
VERSION AF269224.1 GI:8926216
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 4045)
AUTHORS Roulet,A.C., Puel,O., Gesta,S., Alvinerie,M. and Pineau,T.J.
TITLE Analysis of dog MDRI p-glycoprotein
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4045)
AUTHORS Roulet,A.C., Puel,O., Gesta,S., Alvinerie,M. and Pineau,T.J.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-2000) Pharmacology, INRA, 180 Chemin de
Tournesfeuille, BP3, Toulouse 31931, France
FEATURES
source
1..4045
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="original mRNA isolated from Beagle dog colon"
101..3964
/codon_start=1
/transl_table=11
/product="his-tagged-multidrug resistance glycoprotein
MDRI"
/protein_id="AAF01747.1"
/db_xref="GI:8926217"
/translation="MHHHHHDPGKGSAAKNFWMKMGKSKKKKKKKPTVSTFA
MFYSNNLDRLYMLVGTMAAI IHGAALPLMMLVFGNMTDSFANAGISRNKTFVINE
SITNTQHFINHLEEMTYYIYSGIGAVLVAAYIQVSWFCLAAAGRLKIRKQFF

HAIMRQIEGWFDVHVGELNRLTDDVSKINEGIDCKIMGFQSIATFFTFGFIQVGR
GWLKTLVLAISPVLGLSAAIWAIIKLSSTDKELLAVAKAGAAVEVLAARIVIAFG
GOKKELRYNKNLEAKRIGIKKAITANI SIGAIFLLIYASYALAFWYCSLVSSEY
SIGOVLTFEVSILGAFSIGOASPSIEAFANAGAAEYFKIIDNPSIDSYSKSHK
PDNKGLEPNKHFSPSRKEVKILKGLNLKVSQGTVALVNGSGGKSTTVOLMQR
LYDPTGMVCIDGODIETINVRHLREITGVVSEPELVFATTIAENIRYGRNVTMDEI
EKAVKEANAYDFIMKLPNKEDTLVGERAQLSGQKORIAIARALVRNPKILLDEAT
SALDTESEAYVOVALDKARKGRITIVIAHRLSTVRNADVIAGDDGVIVBKGNHDEL
KEGIGYFKLVMTORNEIELEENATGSKSESDALEMPKDSGSLIKRSTRSIIHA
POGDRKLTKEDLNENVPVSEWRILKLNSTWPFYVVGIFCAILINGLOPAPSIIF
SRIGITFREDETPETKQNMFSVLFLVGLIISFTIFFLQGGTFGRAGELLTKRLRY
MVFMSLRQDVSFDPKNTGATLTRLANDAAQVKAIGSLAVLTONTIANLGTGII
ISLYIWLQLLVLAIVPIIAIAGVEMKMLSGQALKKKELEGACKIAFEALENFT
VSLTRGEQYEAQVLOVYRNSLRKAHIFGVFSITOAMHVFVSGFRGAYLV
ANEPMFODVLLVFSATVFCAMAVGOVSPADYAKAKVSAAHVIMLEKSPILDSYS
PHGLKPNLTENGVTFNEVFNYPTRDIPVLOGLSLEVKGQFIALVGGSGCKSVV
QLLERFYDPLAGSVLIDGKELHNLNVOMLRAHLGIQVSEPIIFDCSIAENIATYDMSR
VVSHEEIMQAKENIHHFETIPEKYNTRVGDGTQLSGGOKORIAIARALYRQPHI
LLDDEATSAIDTESEKVVQEAIDKAREGRTCIVIAHRLSTIQNADLIVFPQNGKVEH
GTHOOLLAQRGIYFMSLSVQAGAKR"

BASE COUNT 1171 a 850 c 986 g 1038 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4045
Score: 6454.00 Matches: 1277
Percent Similarity: 99.84% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 2
Query Match: 99.71% Indels: 0
DB: 12 Gaps: 0

US-09-672-725c-25 (1-1281) x AF269224 (1-4045)

QY 2 AspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLysLys 21
DB 122 GATCCTGAAGAGGCCGTGAAGGGAGTGCAGAGAAGAACTTCGGAAAAATGGGCACAAAA 181
QY 22 SerLysLysLysGluLysLysGluLysProThrValSerThrPheAlaMetPheArg 41
DB 182 AGTAAAAAATGAGAGAAGAAAGAAACCAACTGTCAGCAGCTTTGGCAATGTTTCGC 241
QY 42 TyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIleHis 61
DB 242 TATTCAAAATGGCTTGATAGTTGTATATGTTGGTGGGACAATGGCTGCATCATCAT 301
QY 62 GlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAsn 81
DB 302 GGAGCTGCACCTCCCTCATGATGCTGCTTTTGGAAACATGACAGATAGCTTTGCCAAAT 361
QY 82 AlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsn 101
DB 362 GCAGGAATTTCAAGAACAAACACTTTTCCAGTTATATATTAATGAAATATTACGAACAT 421
QY 102 ThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrSer 121
DB 422 ACACAACATTTCAATCAACCATCTGGAGAGGAAATGACCACGATGCTTATTTATACAGT 481
QY 122 GlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeuAla 141
DB 482 GGGATCGGTGCTGGCTGGTGGCTGTACATCCAGGTTTCATTCTGGTGCCTGGCA 541
QY 142 AlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGlu 161
DB 542 GCAGGAAGACAGATACTCAAAATTAGAAAACAATTTTTCATGCTATCATGCCAGAGGAG 601
QY 162 IleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAspVal 181
DB 602 ATTTGGCTGGTTGACGTGCATGACGTTGGGAGGCTTAACACCCGGCTTCACAGACGATGTC 661
QY 182 SerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAlaThr 201
DB 662 TCCAAAATCAATGAAGAAATGGCGACAAAATTTGGAATGTTCTTTCAATCAATAGCAACA 721

QY 202 PhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIleLeu 221
DB 722 TTTTTCACCCGGTTTATAGTGGGGTTTACAGTGGTGGAGGCTAACCCCTTGCAITTTG 781
QY 222 AlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPhe 241
DB 782 GCCATCAGCCCTGTTCTTGACATTTAGCGGCCATCTGGGCAAGATACTATCTTCATTT 841
QY 242 ThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeuAla 261
DB 842 ACTGATAAAGAACTCTTTGGGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGCTTAGCA 901
QY 262 AlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsnLys 281
DB 902 GCATCAGAAGCTGATGGCTTTGGAGGACAAAAGAAAGAACTTGAAGGTACACAAA 961
QY 282 AsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSerIle 301
DB 962 AATTTAGAAGAAGCTAAAGAAATTTGGGATAAAGAAAGCTATCAGCGCAACATTTCTATT 1021
QY 302 GlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSer 321
DB 1022 GGTCCCGCTTCTTATTTGATCTATGCATCATATGCTCTGGCTTCTGGTATGGACCTCC 1081
QY 322 LeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerValLeu 341
DB 1082 TTGGTCTCTCCAGTGAATATTTATTTGGACAAAGTACTCAGCTCTCTTCTTCTGTATT 1141
QY 342 IleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArg 361
DB 1142 ATTTGGGCTTTTATTTGGACAGGCATCCCAAGCATTTGAAGCATTTGCCAAAGCAAGA 1201
QY 362 GlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSer 381
DB 1202 GAGCAGCTTATGAAATCTTCAAGATAATGCAATAAACCAGCAATTCACAGCTATTTCG 1261
QY 382 LysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPhe 401
DB 1262 AAGAGTGGACATAAACACATAATATTAAGGNAATTTGGAATTCAAAATGTTCACTTC 1321
QY 402 SerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSer 421
DB 1322 AGTTACCTCTCTCGAAAGAAGTTAAGATCTTAAAGGGCTCAACCTGAAGGTTCAGAGT 1381
QY 422 GlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeu 441
DB 1382 GGGCACACAGTGGCGCTGGTTGGGAACAGTGGCTGCGGGAAGACGACGACCTGCGAGCTG 1441
QY 442 MetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArg 461
DB 1442 ATGCAGAGCTCTATGACCCACAGATGGCATGGTCTGTATTGTGAGCAGACACATTAGG 1501
QY 462 ThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeu 481
DB 1502 ACCATAATGTAAGGCATCTCGGGAAATTAAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1561
QY 482 PheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAspGlu 501
DB 1562 TTTGGCCACCACCATAGCTGAAAACATTCGCTATGCGCGCGGAAAATGTCACCATGATGAG 1621
QY 502 IleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsnLys 521
DB 1622 ATTGAGAAAGCTGTTAAGGAAGCCAAATGCTATGATTTTATCATGAACTACTAATAA 1681
QY 522 PheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArgIle 541
DB 1682 TTTTGACACTCTGGTTGGAGAGAGGGGCCAGCTGAGTGGTGGGACAGAAAACAGAGAATC 1741
QY 542 AlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThrSer 561
DB 1742 GCCATTGCTCGGGCCCTGGTTGCAACCCCAAGATTTCTTCTGCTGGATGAGCAACGTC 1801
QY 562 AlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLysGly 581

|||||
Db 1802 GCTCTGGACACGTGAAGTGAAGCAGTGGTTTCAGGTGGCCCTGGATAAGGCCCAAGAAAGGC 1861
QY ArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla 601
Db 1862 CGGACTACCATTTGTGATAGCTCATCGTTTGTCTACAGTTTCGTAATGCCGATGTCATTGCT 1921
QY GlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGluLys 621
Db 1922 GGTTTTGATGATGGAGTCATTGTGGAGAAGGAATCATGATGAATCATGAAAGAGAG 1981
QY GlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGluAsn 641
Db 1982 GGCATTTCCTCAAACTTGTCAATGCAGACAAGAGGAATGAAATGAGTTAGTAATAAT 2041
QY AlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSerGly 661
Db 2042 GCCACTGGTGAATCCAAAAGTGAAGTGTATGCTTGGAAATGCTCCAAAAGATTACAGG 2101
QY SerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAsp 681
Db 2102 TCCAGTTTATATAAAGAGATCACTCCGAGGATATACATGCACCACAAGGCCAAGAC 2161
QY ArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPheTrpArg 701
Db 2162 AGAAGCTTGGTACAAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCTCTGGAGG 2221
QY IleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAlaIle 721
Db 2222 ATTTCTGAAGCTGAACCAACTGAATGGCCTTATTTTGTGGTGGTATATTTTGTGCTATT 2281
QY IleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePhe 741
Db 2282 ATAAACGGAGCCCTGCACACGATTTTCAATAATATTTTCAAGGATATAGGATCTTT 2341
QY ThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe 761
Db 2342 ACCCGAGATGAGGATCCTGAACAAAAACGACAGATAGTAACATGTTTCTGTATTGTTT 2401
QY LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys 781
Db 2402 CTAGTCTTGGAAATATTTCTTTTATATATTTTCTCCAGGGCTTCACATTTGGCAAA 2461
QY AlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGln 801
Db 2462 GCTGGGAGATCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTGAGACAG 2521
QY AspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla 821
Db 2522 GATGTCAGCTGGTTGATGACCCCTAAAAACACCACTGGAGCATTTGACAAACAGGCTTGCC 2581
QY AsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn 841
Db 2582 AATGATCGCGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGTCTGTCATTACCCAGAA 2641
QY IleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeu 861
Db 2642 ATAGCAATCTTGGACAGCGCATTTATATATCTTAATCTATGTTGGCAATTAACATT 2701
QY LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMetLeu 881
Db 2702 TTACTCTTAGCAATTGTACCATTCAATGCAATAGCAGAGTGTGTTGAAATGAAATGTTG 2761
QY SerGlyGlnAlaLeuLysAspLysLysGluLeuGluValGlyValGlyLysIleAlaThrGlu 901
Db 2762 TCTGGACAACACTGAAAGATAGAAGAGCTAGAAGAGCTGGGAAGATGCTACAGAA 2821
QY AlaIleGluAsnPheArgThrValSerLeuThrArgGluGlnLysPheGluTyrMet 921
Db 2822 GCCATCGAAAACCTCCGAACCTGTTGTTCTTTGACTCGGAGCAGAGTTTGAATACATG 2881
QY TyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGly 941
|||||

RESULT 6
AF045016
LOCUS

4317 bp mRNA linear

MAM 07-FEB-1998

Db 2882 TATCCACAGAGTTTGAAGTACCATACAGAAACTCTTTTGGAGAAAGCACACATCTTCGGG 2941
QY ValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPhe 961
Db 2942 GTCTCATTTTCTATCACCCAGGCAATGATGATTTTTTCTTATGCTGGCTCTTCCGGTTT 3001
QY GlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSer 981
Db 3002 GGTGCTACTTGTGTGGCAAAATGAGTTTCATGAACTTTCAGGATGTTCTTTGGTATTCTCA 3061
QY AlaIleValPheGlyAlaMetAlaValIcGlnValSerSerPheAlaProAspTyrAla 1001
Db 3062 GCTATTGCTTTGTGGTCATGGCAGTGGCAGCTAGTTTCATTTCTCTCTGACTATGACC 3121
QY LysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIleAsp 1021
Db 3122 AAAGCCAAAGTATCAGCAGCCCGTCATCATGATCATTTGAAAAAGCCCTCTGATTGAC 3181
QY SerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGlu 1041
Db 3182 AGCTACAGCCCTCACGGCCTCAAGCCAAATACGTTGGAAGAAATGTGACATTTAATGAG 3241
QY ValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeuGlu 1061
Db 3242 GTCGTGTTTCAACTATCCCACTCGACACACATCCCGTCTCCAGGGCTGAGCCCTCGAG 3301
QY VallLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrVal 1081
Db 3302 CTGAAGAGGCCAGACGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGAGCACAGTT 3361
QY ValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGlyLys 1101
Db 3362 GTTCAGCTCCTTAGAGCGCTTCTATGACCCCTTGCTTCAGTGCCTAAATGATGGCAAA 3421
QY GluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGlu 1121
Db 3422 GAGATAAAGCACCTGAATGTCCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAGGAG 3481
QY ProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgVal 1141
Db 3482 CCCATCTCTGTTGACTGCAGCATTCGCCGAGAACATTGCCTATGGAGACAACAGCCGGTC 3541
QY ValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisPheIleGlu 1161
Db 3542 GTATCACATGAAGAGATTATCAGCGACGCCAAGGAGGCCAACATACACCATTTCATCGAG 3601
QY ThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGly 1181
Db 3602 ACATCCCTGAGAAATACACACAGAGTAGGACACAAGGAAACCCAGCTCTCTGGTGGC 3661
QY GlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeu 1201
Db 3662 CAGAAACAGCGCATTTGCCATAGCTCGCGCTCTTGTGTAGACAGCCTCATATTTTCTTTG 3721
QY AspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeuAsp 1221
Db 3722 GATGAAGCTACATCAGCTCTGATACAGAAAGTGAAGAGTTGTCCAAAGAGCCCTGGAC 3781
QY LysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsn 1241
Db 3782 AAAGCCAGAGAGGCCGCCACCTGCTGATGCCACCCGCTTGTCCACCATCCAGAA 3841
QY AlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGlnGln 1261
Db 3842 GCAGATTTAATAGTGTGTTTTCAGAAATGGCAAGTCAAGAGGAGCATGGCACATCAACAG 3901
QY LeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLysArg 1281
Db 3902 CTGCTGCTCAGAAAGGCATCTATTTTCCATGATCAGTGTCCAGGCTGGAGAAAAACGC 3961
|||||

DEFINITION Canis familiaris multidrug resistance p-glycoprotein (MDR1) mRNA, complete cds.
ACCESSION AF045016
VERSION AF045016.1 GI:2852440
KEYWORDS dog.
SOURCE
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 4317)
Puel,O., Lepage,J.F., Alvinerie,M., Galtier,P. and Pineau,T.
Direct Submission
TITLE Submitted (28-JAN-1998) Pharmacology, INRA, BP 3, 180 Chemin de Tournesfeuille, Toulouse Cedex 9 31931, France
JOURNAL
FEATURES
Location/Qualifiers
source
1..4317
/organism="Canis familiaris"
/db_xref="taxon:9615"
/tissue_type="colon"
/note="beagle"
1..4317
/gene="MDR1"
70..3912
/gene="MDR1"
/codon_start=1
/product="multidrug resistance p-glycoprotein"
/protein_id="AAC02113.1"
/db_xref="GI:2852441"
/translation="MDPEGGRKSAEKNFWMGKSKKKEKKKPKVTFAMPKRSNW LDRYMLVGTMAAIIHGAAALPLMLVFGNMTDSFANAGISRNKTFPIVINESITNNTQ HFTHLEEMETVAYYSIGTACGLVAAYIQVSWCLAAAGROILKIRKQFFHAIMROE IGFWDHVDGELNRTLDDVSGKINEGIDKGVGFOSIATFTFTGTVGTGPGWKLTLY ILAISPVLGSLRAIMWAKILSSFTDKELLAYAKAGAAVEVLAAIRIVAFGQKKELE RYNNLEEKRIKIKAITANSIGAAFLIITAYALAFWIGTSLVLSSEYITIGQVLT VFSVLGAFSGQASPSIEAFANARGAAEYEIFIIDNKPISIDSYSGKHGPDNIGN LEFNHVSFSPKSEKILKLNKQSGQTVALVNGSGCKSTTVQLMQRLYDPTDG MVDIGQDINTINRHLEITGVVSOEPLVFAITIAENIRYGRNVTMDIEIKAVKEA NAYDFTMKLPDFTLVGERLSCGQKORIAALRNPNKILLDRDETSALDPES EAVQVALDKARGRTTIVIAHRLSTVRNADVTAGPDGVIVKGNHDELMKEKGIY KLYMTQTRNEIELENATGESKESDALEMSPKDSGSLIKRSTRSRTHAQQGRK LGTKDLNPNVSPFWRIKLKLNSTEWPFYVGFICAILNGGLQPAFSIIFSRIGIF TRDEPDTKRNQSMFSLVLGIIISFTFFLQGFQKAGEIILTKRLRYVFRSML RQVDFNDPNTGALTTLRLANDAQQVKGAISRLAVITQNIANLGTGIIISLYGW QTLFLLLAIVPIIAIAGVVMKMLSGQALKDKKEAGKIAATEAENFTVVSLETR EKPQYVYASQLOVPYRNSLRKAHIFGVSPSIQAMVFSYACFREGAYLVANEPNF QDVLVAFSAIVFAMAVQVSSFPADYAKAKYSAAHVIMIIEKSPILDSYSPHGLKPN TLEGNVTNEVFNPTREDIPVLQGLSLEVRKQTLALVSGSGCKSTTVOLLERFY DPLAGSLVDGKEIKHLNWLRAHLGIVSQEPIFLDCSIAENIAYGDNRSVVSHEE MOAAKEANTHHFTETLPKYNTRVGDGQTLQSGQKORIAALRNPNKILLDRDEAT SALDTSEKRVQBALDKAREGRTCVIAHRLSTIQNADLIIVVFQNGKVEHGHQQLL AOKGIYFMSISVQAGAKR"

BASE COUNT 1293 a 844 c 1019 g 1161 t
ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4317
Score: 6425.50 Matches: 1273
Percent Similarity: 99.69% Conservative: 4
Best Local Similarity: 99.38% Mismatches: 3
Query Match: 99.27% Indels: 1
DB: 4 Gaps: 1

US-09-672-725C-25 (1-1281) x AF045016 (1-4317)

QY 1 MetaspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
|||||
Db 70 ATGGATCCTGAAGGAGCGCGTAGGGAGTGCAGAGAAGAACTCTCGAAATGGGCAA 129
QY 21 LysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
|||||
Db 130 AAAAGT---AAAAAGAGAGAAGAAAACCAACTCTCAGACGTTTGAATGTTT 186
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 60
|||||

Db 187 CGCTATTCAATTTGGCTTGATAGTTGTATATGTTGGTGGGACAAATGGCTGCCATCATC 246
|||||
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
|||||
Db 247 CATGGAGCTGCACCTCCCTCTCATGATGCTGGTTTTTGGAAACATGACAGATAGCTTTGCA 306
|||||
QY 81 AsnAlaGlyLeuSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
|||||
Db 307 AATGCAAGGAATTTCAAGAAACAAAACCTTTCCAGTTATATAATTAATGAAAGTATTACGAAC 366
|||||
QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyrTyr 120
|||||
Db 367 AATACACAACATTTTCATCAACCATCTGGAGGAGGAATGACCACGATGCTTATTTATTAC 426
|||||
QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
|||||
Db 427 AGTGGGATCGGTGCTGCGCTGCTGGTGGCTGTACATCCAGGTTTCATTCGTGGTGCCTG 486
|||||
QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
|||||
Db 487 GCAGCAGGAAGACAGATACTCAAAATTAGAAACAATTTTTCATGCTATCATGCGACAG 546
|||||
QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
|||||
Db 547 GAGATTGGCTGGTTTCACGTGCATGACGTTGGGAGCTTAACACCGGCTCACAGACGAT 606
|||||
QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIleAla 200
|||||
Db 607 GTCTCAAAATCAATGAAGAAATGGCGACAAGTTGGAATGTTCTTTCAATCAATAGCA 666
|||||
QY 201 ThrPhePheThrChlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
|||||
Db 667 ACATTTTTCACCGGTTTATAGTGGGGTTTACACCTGGTTGGAAGCTAACCTTGTGAT 726
|||||
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
|||||
Db 727 TTGGCCATCAGCCCTCTTCTTGACCTTCAGCCGCCATCTGGGCAAGATATCTATCTTCA 786
|||||
QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
|||||
Db 787 TTTACTGTAAAGAACTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 846
|||||
QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysLysLysLysLysLysLys 280
|||||
Db 847 GCAGCAATCAGAACTGTGATGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAAAC 906
|||||
QY 281 LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
|||||
Db 907 AAAAAATTTAGAGAAGCTAAAGAAATGGGATAAAGAAAGCTATCACGCCCAACATTCT 966
|||||
QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
|||||
Db 967 ATGGTGGCGCTTCTTATTGATCTATGATCATATGCTATGCTTCTGGTATGGGACC 1026
|||||
QY 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
|||||
Db 1027 TCTTGGTCCCTCCAGTGAATATATCTATTTGGACAGGTACTCAGTCTCTCTTTCTGTGA 1086
|||||
QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
|||||
Db 1087 TTAATTTGGGCTTTTAGTATTGGACAGGATCCCAAGCATTTGAAGCATTTGCAACGCA 1146
|||||
QY 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
|||||
Db 1147 AGAGGAGCAGCTTATGAAATCTTCAAGATAATGACAATAAACCACCAAGCATTCACAGCTAT 1206
|||||
QY 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
|||||
Db 1207 TCGAAGAGTGGACATAAACACAGATAATTAAGGGAAATTTGGAATTCAAAAATGTTCCAC 1266
|||||
QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
|||||

QY 1141 ValValSerHisGluGluMetGlnAlaAlaLysGluAlaAsnHisPheIle 1160
|||||
Db 3487 GTCTATCATGACAGAGATTATGAGGAGCCAGGAGGCCAATACATACACCTTCATC 3546
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
|||||
Db 3547 GAGACACTCCCTGAGAAATACAAACACAGAGTAGGAGACAAAGGAAACCCAGCTCTGTGT 3606
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
Db 3607 GCCCAGAAACAGCGATTGCCATAGCTCGCGCTCTGTGTAGACAGCCCTCATATTTGTGT 3666
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
Db 3667 TTGGATGAAGCTACATCAGCTCTGATACAGAAAGTAGAAAGGTTGTCCAGAAGCCCTG 3726
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
Db 3727 GACAAAGCCAGAGAGCGCGCACCTGTCATTGTGATCGGCCACCGCTTGTCCACCATCCAG 3786
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
Db 3787 AATGCAGATTTAATAGTGTGTTTCAGAAATGGCAAGTCAAGGAGCATGGCACACATCAA 3846
QY 1261 GlnLeuLeuAlaGlnLysGlyIleYrPheSerMetValSerValGlnAlaGlyAlaLys 1280
Db 3847 CAGCTGCTGGCTCAGAAAGGCATCTATTTTCCATGATCAGTGTCCAGGCTGGAGCAAG 3906
QY 1281 Arg 1281
|||
Db 3907 CGC 3909

RESULT 7
AX105059
LOCUS
DEFINITION Sequence 3 from Patent WO0123540.
ACCESSION AX105059
VERSION AX105059.1 GI:13921211
KEYWORDS
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 4317)
AUTHORS Stocker,P.J., Steimel-crespi,D.T., Crespi,C.L., Reif,T.C. and Patten,C.J.
P-glycoproteins and uses thereof
TITLE Patent: WO 0123540-A 3 05-APR-2001;
JOURNAL CORPORATION (US)
FEATURES
Location/Qualifiers
source
1..4317
/organism="Canis familiaris"
/db_xref="taxon:9615"
70..3912
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37719.1"
/db_xref="GI:13921212".
/translation="MDPEGRKSGAEKNFWMKKGSKKEKKPTVSTFAMFRYSNN
LDRLYMLVGTMAAITHGAALPLMLVFNMTDSFANAGISRNKTFPVILINESITNTQ
HFINHLEEMTYAYYSGIGAGLVAAAYIOVSFWCLAAAGROILKIRKQFPAIMROE
IGWPDVDVDELNRLTDVSKINEGDKVGMFPQSIATFTFTGFIYVGTTPGWKLTIV
ILASPVGLSNAWAKILSSFTDKELLAYAKAGAAVEVLAAITRTVAFGGQKKELE
RYNKLEAKRINKKATANISIAAFLLIYASTALAFWGTSLVLSSETTIGQVLT
VFFSVLIGAFSIGQASPISEAFANAGAAEYFKIIDNKPISIDYSKSGKPDNIKGN
LEFNKMHFSYSPSRKEVKILGKLNKLVQSGQTVALVNGCGCKSTVOLMQRLYDPTDG
MYCIDGDIRITNVRLHREITGVVSOEVLPAFTIAENIRYGRNVMTDETEKAVKEA
NAYDFIMKLNPKEFTLVGRCARGSKGQKQRIARALVRNPKIILLDEATSAVDTES
EAVQVLDKARKKRTTIVIAHRLSTVRNADVIAGFDGQVIVEKGNHDELMEKGIITP
KLVTQMTRNGNEIELENATGESKSESDALEMSPKDSGLIKRRSTRKRSIHAPOQDHRK
LGTREDLENVPSYFWRLKLNSTEWPFVYVFGICAILNGLOPAFSIIFSRIGIF
TRDEPETKRONSNMFSVLFLVGIISFTFLOGFTFGKAGEILITKRLRYWVFRSML
RDQVSNFDDPNKTTCALTTRLANDAAQVKGAGISFLAVITQNIANLGTGIIISLYGW

BASE COUNT 1293 a 844 c 1019 g 1161 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4317
Score: 6425.50 Matches: 1273
Percent Similarity: 99.69% Conservative: 4
Best Local Similarity: 99.38% Mismatches: 3
Query Match: 99.27% Indels: 1
DB: 6 Gaps: 1
US-09-672-725C-25 (1-1281) x AX105059 (1-4317)
QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
|||||
Db 70 ATGGATCCTGAAGGAGCGCGTAAGGGAGTGCAGAGAGAACTTCTGGAAATGGGCAA 129
QY 21 LysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
|||||
Db 130 AAAAGT---AAAAAGAGAGAGAAAGAAACCACTGTCCACGACGTTTGGCAATGTTT 186
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 60
|||||
Db 187 CGCTATTCAAATTGGCTTGATAGCTGTATATGTTGGTGGGACAATGGCTGCCATCATC 246
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
|||||
Db 247 CATGAGCTGCACCTCCCTCTCATGATGCTGTTTGGAAACATGACAGATAGCTTTGCA 306
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 307 AATCAGGAATTTCAAGAAACAAACTTTCCAGTTTATATTAATGAAGATATTACGAAC 366
QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 120
Db 367 AATACACAACTTTCATCAACCATCTGAGGAGGAAATGACCACGTATGCCTATTATTAC 426
QY 121 SerGlyIleGlyAlaGlyValLeuValAlaIleValSerPheTrpCysLeu 140
Db 427 AGTGGGATCGGTGCTGCTGCTGCTTACATCCAGGTTTCATTCTGTGCTGTG 486
QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
Db 487 GCAGCAGGAAGACAGATACTCAAAATTAGAAAACAATTTTTCATGCTATCATCGGACAG 546
QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
Db 547 GAGATTGCTGGTTTGACGTGCATGACGTTGGGAGCTTAACACCCGGCTCACAGACGAT 606
QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
Db 607 GTCTCCAAAATCAATGAAGGAATTTGGGCACAAAGTTTGAATGTCTTTCATCAATAGCA 666
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
Db 667 ACATTTTTCACCGGTTTATAGTGGGTTTACACTGGTGGAGACGTAAACCTTGTGAT 726
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
Db 727 TTGGCCATCACCCCTGTTCTTGGACTTTCAGCCGCATCTGGGCAAGAGATATCTTCA 786
QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 260
Db 787 TTTACTGATAAAGAACTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 846

QY 261 AlaAlaIleAthrGhrValIleAlaIlePheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
DB 847 GCAGCAATCAGAACTGTGATTGCTTTGGAGGACAAAAGAAAGAACTTGAAGGTTACAAC 906
QY 281 LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
DB 907 AAAAATTTAGAAGAGCTAAAGAAATTTGGGATTAAGAAAGCTATCACGGCCCAACATTTCT 966
QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrTyrGlyThr 320
DB 967 ATTTGGTCCCTTCTTATTGATCTATGCATCATATGCTTGGCTTCTGGTATGGGACC 1026
QY 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
DB 1027 TCTTGGTCTCTCCAGTGAATATCTATTGGACAGGTACTCAGTCTCTCTCTTTCTGTA 1086
QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
DB 1087 TTAATTGGGGCTTTTAGTATTGGACAGGCATCCCAAGCATTTGAAGCATTTGCCAAACGCA 1146
QY 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
DB 1147 AGAGGACAGCTTATGAATCTTCAAGATAATTGACAATAAACCAAGCATTTGACAGCTAT 1206
QY 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
DB 1207 TCGAAGAGTGGACATAAACCCAGATAATATTAAAGGAAATTTGGAAATTCAAAAATGTTTCAC 1266
QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
DB 1267 TTCAGTTACCTTCTCGAANAAGATTAAGATCTTAAGAGGCTCAACCTGAAGGTTTCAG 1326
QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
DB 1327 AGTGGGCAGACAGTGGCGCTGGTGGGAACAGTGGCTGGGGGAAGACAGCCGCTGCAG 1386
QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
DB 1387 CTGATCGAGAGGCTCTATGACCCACAGATGGCATGGTCTGATTTGATGACAGGAGCAT 1446
QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
DB 1447 AGACCAATAAATGTAGGCATCTTCGGGAATTTACTGGTGGTGGTGGTGGTGGTGGTGG 1506
QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
DB 1507 TTTGTTGCCACCACGATAGCTGAAACATTCGCTATGGCCGCGAAAAATGTCACCATGGAT 1566
QY 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
DB 1567 GAGATTGAGAAAGCTGTTAAGGAAGCCAAATGGCTATGATTTTATCATGAACACTACCTAAT 1626
QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
DB 1627 AAATTTGACACTCTGGTGGAGAGAGGGGGCCGCTGAGTGGTGACAGAGAAACAGAGA 1686
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
DB 1687 ATGCCATTGCTCGGGCCCTGGTTGCGCAACCCCAAGATCTTCTGCTGGATGAGGCAAG 1746
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
DB 1747 TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGTGGCCCTGGATAAGGCCAGAAAA 1806
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
DB 1807 GCCCGGACTACCATTTGATAGCTCATGCTTGTCTACAGTTTCGTAATCCCGATGTCATT 1866
QY 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
DB 1867 GCTGGTTTGTATGATGAGTCAATTTGGAGAAAGGAATCATGATGAACATCATGAAAGAG 1926
QY 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640

DB 1927 AAGGCACTTTACTTCAAATCTGTCAATGTCAGACAGAGGAAATGAAATTGAGTTAGAA 1986
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
DB 1987 ATNGCCACTGGTGAATCCAAAAGTGAAGTATGCTTGGAAATGCTCCCAAAAGATCA 2046
QY 661 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
DB 2047 GGGTCCAGTTTAATAAAAAGAAGATCAACTCGCAGGAGTATACATGCACCAAGGCCAA 2106
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTyr 700
DB 2107 GACGAAAAGCTTGGTACAAAAGAGACTTGAATGAGAAATGTTCTTCCAGTTTCTCTGG 2166
QY 701 ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 720
DB 2167 AGGATCTCGAAGCTGAACCTCAACTGAATGGCTTATTTTGGTGGTGGTATATTTGGCT 2226
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
DB 2227 ATTAATAACGGAGGCTGCAACACAGCATTTTCAATAATATTTTCAAGGATATAGGGATC 2286
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
DB 2287 TTTATCCCGAGATGAGGATCCTGAAACAAAACGACAGAAATAGTAACATGTTTCTGTATTG 2346
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
DB 2347 TTTCTAGTCCCTGGATTAATTTCTTTTATTACATTTTCTCCAGGGCTTCACATTGGC 2406
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
DB 2407 AAAGCTGGGAGATCCTCACTAAGCGGCTTCGATACATGGTTTTCAGATCCATCCTGAGA 2466
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
DB 2467 CAGGATGTACGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2526
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
DB 2527 GCCAATGATCGGCTCAAGTTAAAGGGCTATAGTTCAGGCTTGTGTCTATACCCAG 2586
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTyrGlnLeuThr 860
DB 2587 AATATAGCAAAATCTTGGGACAGCATTAATATATCTTAATCTATGTTGGCAATTAACA 2646
QY 861 LeuLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLysMet 880
DB 2647 CTTTACTCTTAGCAATTTGTACCCATCATTCGAATAGCAGGAGTTGTGAATGAAAATG 2706
QY 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
DB 2707 TTGTCGACAAAGCAGCTGAAAGATAAGAAAGAGCTAGAAAGAGCTGGGAAGATGCTACA 2766
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
DB 2767 GAAGCAATCGAAAAGCTCCGAACTTCTGTTTCTTTGACTCGGGAGCAGAAAGTTTGAATAC 2826
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
DB 2827 ATGTATGCACAGATTTGCAAGTAGCCATACAGAACTCTTTGAGGAAAGCACACATCTTC 2886
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
DB 2887 GGGGTCTCATTTTCTATCACCCAGCAATGATGATTTTCTTCTATGCTGGCTGTTCCGG 2946
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
DB 2947 TTTGGTGGCTACTTGGTGGCAAAATGAGTTTCATGAACCTTCAGGATGTTCTTTTGGTATTC 3006
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000

Db	3007	TCAGCTATTGCTCTTTGGTGGCCATGGCAGTGGCCAGGTGAGTTCAGTTCTCTGCTGACTAT	3066
QY	1001	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleLysSerProLeuLeu	1020
Db	3067	GCCAAAGCCAAAGATATCAGCAGCCACGTCATCATGATCATTTGAAAGAGCCCTCTGATT	3126
QY	1021	AspSerTySerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1040
Db	3127	GACAGTACAGCCCTCAGCGCCCTCAAGCCAAATACCTTGAAGAAATGTGACATTTAAT	3186
QY	1041	GluValValPheAsnTyProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1060
Db	3187	GAGTCTGTGTCAACTATCCACTCCAGCAGACATCCCGTGTCTCCAGGGGTGACCTC	3246
QY	1061	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1080
Db	3247	GAGGTGAAGAAGGCCAGAGCGCTGGCCCTCGTAGTAGAGTGGCTGTGGGAAGACACA	3306
QY	1081	ValValGlnLeuLeuGluArgPheTyArgProLeuAlaGlySerValLeuIleAspGly	1100
Db	3307	GTGTTTCAGCTCTAGAGCGCTTCTATGACCCCTTGCTGGTTCAGTGCTAATGTATGGC	3366
QY	1101	LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1120
Db	3367	AAAGAGATAAGCACCTGAATGTCAGTGGCTCGGAGCACACCTGGCATCGTGTCTCAG	3426
QY	1121	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyArgGlyAspAsnSerArg	1140
Db	3427	GAGCCCATCTCTGTGACTGCAGCATTCGCGAGAATTCGCCATGGAGACAACACCGCG	3486
QY	1141	ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle	1160
Db	3487	GTGCTATCAGATGAAGAGATTATGCAGCGAGCCAGAGGCCAATACACCATCTTCATC	3546
QY	1161	GluThrLeuProGluLysTyArgThrArgValGlyAspLysGlyThrGlnLeuSerGly	1180
Db	3547	GAGACATCTCCTGAGAATAACACACCCAGAGTAGGAGACAAGAACCCAGCTCTCTGGT	3606
QY	1181	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1200
Db	3607	GGCCAGAAACAGGCGCATTTGCATAGCTCGCGCTCTTTAGACAGCGCTCATATTTTGCTT	3666
QY	1201	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1220
Db	3667	TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAAGAGCCCTG	3726
QY	1221	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1240
Db	3727	GACAAAGCCAGAGAGCGCCACCTGCTGATGTCGCCCCACCGCTGTGCCACCATCCAG	3786
QY	1241	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1260
Db	3787	AATGCAGATTTAATAGTGGTGTTCAGAAATGGCAAGTGCAGAGCATGGCACACATCAA	3846
QY	1261	GlnLeuLeuAlaGlnLysGlyIleTyThrPheSerMetValSerValGlnAlaGlyAlaLys	1280
Db	3847	CAGCTGCTGGCTCAGAAGGCGATCTATTTTCCATGATCAGTGTCCAGGCTGGAGCAAG	3906
QY	1281	Arg 1281	
Db	3907	CGC 3909	
RESULT 8			
CFA419568			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Canis familiaris			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
1 (sites)			
Roulet,A			
Characterization of a MDR1a P-gp deficient dogin relation to			
ivermectin sensitivity			
Thesis (2002) Department of Biological Sciences, Universite de			
Toulouse (UPS III), Toulouse, France			
2 (bases 1 to 3934)			
Roulet,A			
Direct Submission			
Submitted (07-NOV-2001) Roulet A., Pharmacology, Inra, 180 Chemin			
de Tournefeuille, 31931 Toulouse cedex, FRANCE			
Location/Qualifiers			
1..3934			
/organism="Canis familiaris"			
/db_xref="taxon:9615"			
/tissue_type="liver"			
/note="ivermectin sensitive collies"			
7..282			
/gene="p-gp"			
7..282			
/gene="p-gp"			
/note="abortive product"			
/codon_start=1			
/evidence="experimental"			
/product="P-glycoprotein"			
/protein_id="CAD12806.1"			
/db_xref="GI:17385393"			
/translation="MDPEGRRGSAEKNFWMGKSKKNEKKGKPTVSTFAMFRYSN			
WDLRLMLVGTMAAIIHGAALPLMLVFGNLTALQMOEFQETKLFQL"			
BASE COUNT			
ORIGIN			
1144 a 815 c 964 g 1011 t			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
US-09-672-725C-25 (1-1281) x CFA419568 (1-3934)			
QY	1	MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys	20
Db	7	ATGGATCTCTGAGGAGCGCTAGGGGAGTGCAGAGAAGAACTTCTGGAAAGATGGCAAA	66
QY	21	LysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe	40
Db	67	AAAAGTAAAAAATCAGAGAAGAAAGAAACCAACTGTCAGCAGCTTTCGAATGTTT	126
QY	41	ArgTySerAsnTrpLeuAspArgLeuTyMetLeuValGlyThrMetAlaAlaIleIle	60
Db	127	CGCTATTCAAAATGGCTTATGATGTTGTATATGTTGGTGGGACAAATGGCTGCCATC	186
QY	61	HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	80
Db	187	CATGGAGCTGCACTCCCTCTCATGATGCTGTTTGGAAACAT----GACAGCTTTGCA	242
QY	81	AsnAlaGlyLysSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	100
Db	243	AATGCAGGAATTTCAAGAAACAAACCTTTTCCAGTTATATAATTAAGAAAGTATTAG	302
QY	101	AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyAlaTyTyTy	120
Db	303	AATACACACATTTTCAACACCATCTGGAGGAGAAATGCCACGATATGCTATTATTAC	362
QY	121	SerGlyIleGlyAlaGlyValLeuValAlaAlaTyIleGlnValSerPheTrpCysLeu	140
Db	363	AGTGGATCGGTGCTGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	422
QY	141	AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln	160
Db	423	GCAGCAGGAGACAGATCTCAAAATTAAGAAACAAATTTTTCATCTATCATGCGACAG	482

QY	161	GlulIeGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp	180
DB	483	GAGATTGGCTGGTTGACGTGCATACGTGGGAGCTTAAACACCGCGTCAACAGCAT	542
QY	181	ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla	200
DB	543	GTCTCCAAATCATGAAGAGATCGCGGACAAATTTGGAATGTCTTCAATCAATAGCA	602
QY	201	ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle	220
DB	603	ACATTTTTCACCGGTTTTATAGTGGGTTTACACGTGGTTGGAAGCTAACCCCTTGTGATT	662
QY	221	LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer	240
DB	663	TTGGCCATCAGCCCTGTCCTTGACATTCAGCGCCCATCTGGGCAAAAGATACTACTTCA	722
QY	241	PheThrAspLysGluLeuLeuAlaTyraAlaLysAlaGlyAlaValAlaGluGluValLeu	260
DB	723	TTTACTGATAAAGACTCTTGGCCTATGCANAAGCTGGAGCACTAGCTGAAGAAGTCCTTA	782
QY	261	AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn	280
DB	783	GCACAAATCAAGACTGTGATTGGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAAC	842
QY	281	LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer	300
DB	843	AAAAATTTGAAGAAGCTAAAGAAATTTGGGATAAAGAAAGCTATCACGCCCAACATTCT	902
QY	301	IleGlyAlaAlaPheLeuLeuIleTyraAlaSerTyraAlaLeuAlaPheTrpTyrGlyThr	320
DB	903	ATTGGTCCGCTTCTATTGATCATCATATGCTCTGGCTTCTGGTATGGGACC	962
QY	321	SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal	340
DB	963	TCCTTTGGTCTCTCCAGTGAATATTCTATTGGACAAGTACTCACGTCTTCTTTCTGTA	1022
QY	341	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	360
DB	1023	TTAATTGGGCTTTTATTTAGTAGGACAGGCATCCCAAGCATTTGAAGCATTTGCAAGCGCA	1082
QY	361	ArgGlyAlaAlaTyraGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	380
DB	1083	AGAGGACAGCTTATCAAAATCTTCAAGATAATTGACAATAAAGCAAGCATTGACAGCTAT	1142
QY	381	SerLysSerGlyHisLysProAspAsnIleLysGlyLysAsnLeuGluPheLysAsnValHis	400
DB	1143	TCGAAGAGTGGACATAAACACAGATAATATAAGGAAATTTGGAATTCAAAAATTTGCAC	1202
QY	401	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	420
DB	1203	TTCAGTTACCTTCTCGAAAAGAGTTAAGACCTTAAAGGTCCTCAACCTGAAGGTTCCAG	1262
QY	421	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln	440
DB	1263	AGTGGGCACAGCTGGCGCTGGTTGGAAACAGTGGCTCGCGGAAGAGCAGCACCGTCGAG	1322
QY	441	LeuMetClnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	460
DB	1323	CTGATGCAGAGGCTTATGACCCACAGATGGCATGGTCTCTGTTATGTATGGACAGGACATT	1382
QY	461	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	480
DB	1383	AGACCATTAATGTAAAGGCATCTTCGGGAATTTACTGGTGGTGAGTCAGGAGCGGTGTG	1442
QY	481	LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	500
DB	1443	TTGTTTGCACACGATAGCTGAAACATTCGCTATGGCCCGCGAAATGTGCACCATGGAT	1502
QY	501	GluIleGluLysAlaValLysGluAlaAsnAlaTyraAspPheIleMetLysLeuProAsn	520
DB	1503	GAGATTGAGAAGCTGTTAAGGAGCCAAATGCCTATGATTTTATCATCAAACTACCTAAT	1562

QY	521	LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlnLysGlnArg	540
DB	1563	AAATTTTGACACTCTGGTTGGAGAGAGAGGGCCAGCTGAGTGGTGGACAGAAACAGAGA	1622
QY	541	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAlaThr	560
DB	1623	ATCGGCATTCGTGGGGCCCTGGTTGCGAACCCCAAGATTCTTCGTGGTATGAGGCAACG	1682
QY	561	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	580
DB	1683	TCAGCTCGGACACTGAAGTGAAGCAGTGGTTCAGTGGCCCTGGATAAGCCAGAAAA	1742
QY	581	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	600
DB	1743	GGCCGGACTACCATTTGATAGCTCATCGTTTCTACAGTTTCGTAAATGCCGATGTCATT	1802
QY	601	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	620
DB	1803	GCTGTTTGTGATGATGAGTCACTTGGAGAAAGGAAATCATGATGAACCTCATGAAGAG	1862
QY	621	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	640
DB	1863	AAGGCATTTACTTCAAACTGTCCAAATGCACACAGAGGAAATGAATTTGAGTTAGAA	1922
QY	641	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	660
DB	1923	AATGCCACTGGTGAATCCAAAAGTGAAGTGATGCCTTGGAAATGCTCCAAAAGATTCA	1982
QY	661	GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln	680
DB	1983	GGGTCCAGTTTAATAAAGAAAGATCAACTCGCAGGAGTATACATGCACACAGGCCAA	2042
QY	681	AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp	700
DB	2043	GACACAAGCTTGGTACAAAAGAGGACTGGAATGAGAATGTACCTCCACTTTCTCTCGG	2102
QY	701	ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla	720
DB	2103	AGGATTCTGAAGCTGAACCTCAACTGAATGGCCTATTATTTGTGGTGGTATATTTGTGCT	2162
QY	721	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	740
DB	2163	ATTATAACGGAGGCCCTGCACACAGCATTTTCAATAATATTTCAAGGATTATAGGATC	2222
QY	741	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	760
DB	2223	TTTACCCGAGATGAGGATCCTGAACAAACACACAGATAGTAACTGTTTCTGTATTG	2282
QY	761	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	780
DB	2283	TTTCTAGTCCTTGGAAATTATTTCTTTTATACATTTTCTCCAGGGCTTCACATTTGGC	2342
QY	781	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	800
DB	2343	AAAGCTGGGAGATTCCTCACTAAGCGGCTTCGATACATGTTTCAGATCCATCCTCAGA	2402
QY	801	GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	820
DB	2403	CAGGATGTCACTGGTTGATGACCCATAAACACACCATGGAGCATTTGACACCCAGCCTT	2462
QY	821	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	840
DB	2463	GCCAAATGATCGGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGCTGTCATTACCCAG	2522
QY	841	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr	860
DB	2523	AATATAGCAATCTTGGGACAGCATTTATATATCTTAATCTATGTTGGCAATTAACA	2582
QY	861	LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet	880
DB	2583	CTTTTACTCTTAGCAATGTACCATCATTTGCAATAGCAGGAGTTGTGAAATGAAATG	2642
QY	881	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyValAlaGlyLysIleAlaThr	900


```
|||||
Db 2643 TTGCTGGACAAGCAGCTGAAAGATAGAAAGAGCTAGAGGAGCTGGGAAGATTCCTACA 2702
QY GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGluLysPheGluTyr 920
Db 2703 GAAGCCATCGAAACTTCGGAAGCTGTGTTCTTTGACTCGGGAGCAGAAAGTTTGAATAC 2762
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2763 ATGTATGCACAGAGTTGCAAGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTTC 2822
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2823 GGGGCTCATTTTCTATCACCCAGCAATGATGATATTTTCTCTATCTGCTGGCTTTCGG 2882
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
Db 2883 TTTGGTGCTACTTGGTGGCAATGAATCATGAATTTTCAGGATGTTCTTTTGGTATTC 2942
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
Db 2943 TCAGCTATGTCTTGGTGCCATGCGAGTGGGACAGTCAGTTCAATTTGCTCTCGACTAT 3002
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProIleuIle 1020
Db 3003 GCCAAGGCCAAAGTATCAGACGCCACCTCATCATGATCATTTGAAAGAACCCCTCTGATT 3062
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3063 GACAGCTACAGCCCTCAGCGGCTCAAGCCAAATAGTTTGGAGGAAATGTGACATTTAAT 3122
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3123 GAGGTGCTGTCACTATCCACTCGACACAGACATCCCGGTCTCCAGGGGCTCAGCCCTC 3182
QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
Db 3183 GAGGTGAAGAGGGCCACAGCTGSCCTCGTAGGTAGCAGTGGCTGGGGAAGACACA 3242
QY 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
Db 3243 GTTGTTCAGCTCCTAGACGCTTCTATGACCCCTTGGCTGGTTCAGTCTCAATTCATGTC 3302
QY 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120
Db 3303 AAAGAGATAAGCACCTCAATGTCCAGTGGCTCCGAGCACACCTGGGCTGCTGTCTCAG 3362
QY 1121 GluProfileLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
Db 3363 GAGCCCATCTCTTGTACTGCAGCATTCGCCGAGAATTCGCTTATGGAGACAACAGCCGG 3422
QY 1141 ValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisIlePheIle 1160
Db 3423 GTGCTATCACATGATGAAGAGATTATGAGGAGCCCAAGGAGGCCAATACACACCTTCATC 3482
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
Db 3483 GAGACACTCCCTGAGAAATACACACCCAGATAGAGACAAGAACCCAGCTCTCTGGT 3542
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
Db 3543 GCCCAAGAACAGCGCATTCGATAGTCCGCTCTTGTGTAGACAGCTCATATTTTGGCTT 3602
QY 1201 LeuaspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
Db 3603 TTGGATGAAGCTACATCAGTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAGCCCTG 3662
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
Db 3663 GACAAGCCAGAGAGGCCGCACTGCAATTGTGATCGCCCACTGCTGTCACCATCCAG 3722
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
|||||
```

```
Db 3723 AATCAGATTATATAGTGGTGTTCAGAAATGCGAAAGTCAAGGAGCATGGCACACATCAA 3782
QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
Db 3783 CAGCTGCTGGCTCAGAAAGGATCTATTTTCCATGATCAGTGTCCAGGTGGAGCAAAA 3842
QY 1281 Arg 1281
Db 3843 CGC 3845

RESULT 9
AX322787
LOCUS AX322787 3860 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 1 from Patent WO0192877.
ACCESSION AX322787
VERSION AX322787.1 GI:18093766
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Sorrentino,B. and Schuetz,J.
TITLE Method of identifying and/or isolating stem cells
JOURNAL Patent: WO 0192877-A 1 06-DEC-2001;
ST. JUDE CHILDREN'S RESEARCH HOSPITAL (US)
FEATURES
Location/Qualifiers
source
I. 3860
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1135 a 746 c 957 g 1022 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3860
Score: 5866.00 Matches: 1164
Percent Similarity: 95.09% Conservative: 55
Best Local Similarity: 90.80% Mismatches: 59
Query Match: 90.62% Indels: 4
DB: 6 Gaps: 4

US-09-672-725c-25 (1-1281) x AX322787 (1-3860)
QY 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1 ATGATCTTGAAGGGGACCCGAATGGAGGAGCAAGAAAGAACTTTTAAACTGAAC 60
QY 20 LysLysSerLysLysGlyLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 61 AATAAAGT--GAATAAGATAAGAGGAAAGAAACCAACTCTCAGTGTATTTTCAATG 117
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 118 TTTTCGCTATTCAAATTTGGCTTGACAAAGTTGATATGTTGGTGGAACTTTGGTGGCATC 177
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 178 ATCCATGGGGCTGACCTTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 237
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 238 GCAAATGCAGCA--AATTTAGAGATCTGATGTCAAACATCACTAATAGAGTGATATC 294
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 119
Db 295 AATGATACAGGGTCTTCATGAAT---CTGGAGGAAAGACATGACCATGCTATTAT 351
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 352 TACAGTGAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 411
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
|||||
```

Db 412 CTGGCAGCTGGAGACAAATACACAAAATAGAAAACAGTTTTTCATGCTATAATGCGA 471
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 472 CAGGAGATAGGCTGGTTTGATGTGCACGATGTGGGAGCTTAACACCCGACTTACAGAT 531
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db 532 GATGTCCTCTAAGATTAAATGAAGTATTGGTGACAAATTTGGAATGTCTTTCAAGTCAATG 591
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 592 GCACCAATTTTCACTGGGTTTATAGTAGGATTTACACGTGGTTGGAAGCTAACCCCTTGG 651
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 652 ATTTTGGCCATCAGTCTCTTCTTGACGTGCAGCTGCTGCTGGCCAAAGATACTATCT 711
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
Db 712 TCATTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTCAAGAGGTC 771
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysGluLeuGluArgTyr 279
Db 772 TTGGCAGCAATTAGAACTGTGATTGCAATTTGGAGGACAAAAGAAAGAACTTGAAGGTAC 831
QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 832 AACAAAAATTTAGCAAGCTAAAGAAATTTGGATAAAGAAAGCTATTACAGCCAATATT 891
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 892 TCTATAGGTCCTCTTCTGCTGATCTATGCATCTTATGTCTGCGCTTCTGGTATGGG 951
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 952 ACCACCTTGGTCCTCTCAGGGGAAATATCTATTTGGCAAGTACTCAGTGTATCTTTCT 1011
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1012 GTATTAATTTGGGCTTTTAGTGTGGACAGGCATCTCCAAGCATTTGAAGCATTTGCAAT 1071
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1072 GCAAGAGGACGACTTATGAANAATCTCAAGATAAATTTGAATAAGCAAGTATTGACAGC 1131
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1132 TATTCGAAGAGTGGGCACAAACCCAGATAATATTAAGGGAATTTGGAATTCGAATGTT 1191
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1192 CACTTCAGTTACCATCTCGAAAAGAAGTTAAGATCTTGAAGGCGCTGAACCTGAAGGTG 1251
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1252 CAGAGTGGGCAGCGTGGCCCTTGGTTGGAACAGTGGCTGTGGGAAGAGCACAAACAGTC 1311
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1312 CAGCTGATGCGAGGCTCTATGACCCACAGAGGGATGGTCAGTGTGATGGACAGGAT 1371
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1372 ATTAGGACCATAAATGTAAGGTTTCTACGGGAATCATTTGGTGTGGTGAAGCAACCT 1431
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1432 GTATTTGTTGCCACCGATAGCTGAAAACATTTGCGTGTGGCCGTGAAAATGTGCACATG 1491
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1492 GATGAGATTGAGAAAGCTGTCAAGGAAGCAATGCCTATGACTTTTATCATGAAGTCGCT 1551

QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1552 CATAAATTTACACCCCTGGTTGGAGAGAGGGGCCCAAGTTGAGTGGTGGCGAAGACAG 1611
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 1612 AGGATCGCATGTGCACGTGGCCCTGGTTCCGAAACCCCAAGATCTTCCTGCTGGATGAGGCC 1671
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 1672 ACGTTCAGCCTTGGACACAGAAAGCGAAGCAGTGGTTTCAGGTGGCTCTGGATAAGGCCAGA 1731
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 1732 AAAGTCGGACCCACCATTTGTATAGCTCATCGTTTGTCTACAGTTCGTAAATGCTGACGTC 1791
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 1792 ATCCTGCTGGTTTCGATGATGGAGTCATTTGGAGAAAGAAATCATGATGAACATCATGAAA 1851
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 1852 GAGAAAGCATTTACTTCAAACTTGTCCAAATGCAGACAGCAGGAAATGAAGTTGAATTA 1911
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuLeuMetSerProLysAsp 659
Db 1912 GAAAATGCAGCTGATGAATCCAAAAGTGAATTTGATGCTTGGAAATGTCTTCAAAATGAT 1971
QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 1972 TCAAGATCCAGCTCTAATAAGAAAAGATCAACTCGTAGGAGTGTCCCTGGATCACAGCC 2031
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2032 CAAGACAGAAAGCTTAGTACCACAAAGAGCTCTGGATGAAAGTATACCTCCAGTTTCCTTT 2091
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2092 TGGAGGATATTGAAGCTAAATTTAACTGAATGGCTTATTTTGTGTGGTGTATTGTGT 2151
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2152 GCCATTATAATGAGGCGCTCCACACCACTTTCGAATAATATTTTCAAAGATTATAGGG 2211
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2212 GTTTTACAGAATTTGATGATCCTGAAACAAACGACAGATAGTAACTTGTTTTTCACTA 2271
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2272 TTGTTTCTAGCCCTTGGAAATTTATTTCTTTTATTACATTTTCTTCAAGGTTTTCACATTT 2331
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2332 GGCRAAGCTGGAGAGATCCTCACCAGGCGCTCCGATACATGGTTTCCGATCCATGCTC 2391
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2392 AGACAGGATGTGAGTGTGGTTTGTATGACCTTAAAAACACCACTGGAGCATTTGACTTACCAGG 2451
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2452 CTCGCCAATGATGCTCTCAAGTTAAAGGGGCTATAGTTCACAGCTTCTGTTGAATTACC 2511
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2512 CAGATATAGCAAAATCTTGGCAGCAGGAATAATTATATCTTCTATCTATGTTGCAACTA 2571
QY 860 ThrLeuLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLys 879
Db 2572 ACATGTTTACTCTTAGCAATTTGTACCCCATCTATGCAATAGCAGGAGTTGTTTGAATAA 2631

```
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
|||||
Db 2632 ATGTTGTCGGACAAGCACTGAACATAGAAGAACTAGAGGTGCTGGGAAGATCGCT 2691
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
|||||
Db 2692 ACTGAAGCAATAGAAAACCTCCGAACCGTTGTTCTTGACTCAGGAGCAAGTTGAA 2751
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
|||||
Db 2752 CATATGATATCAGAGTTTCAGAGTACCACATACAGAAACTCTTTGAGGAAGACACATC 2811
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 2812 TTTGCAATTACATTTTCTTCCACCAGCAATGATGATATTTTCTATGCTGGATGTTTC 2871
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
|||||
Db 2872 CGGTTTGGAGCGCTACTTGGTGGCACAATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 2931
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
|||||
Db 2932 TTTTCAGCTGTGCTTTGGTGCCATGCGCGTGGGCAAGTCAGTTCATTTGCTCCCTGAC 2991
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
|||||
Db 2992 TATGCCAAGCCAAAATATCAGCAGCCACATCATCATGATCATTTGAAAAACCCCTTG 3051
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
|||||
Db 3052 ATTGACAGCTACAGCAGGAAGGCGCTAATGCCGAACACATTTGGAAGGAAATGTCACATTT 3111
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
|||||
Db 3112 GGTGAAGTTGATTCACATATCCACCGCCGACATCCAGTCCAGTTCAGGGACTGAGC 3171
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyLysSer 1079
|||||
Db 3172 CTGGAGGTGAAGAGGGCCAGACGCTGCTGCTGGGCGACGAGTGGCTGGGAAGAGC 3231
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
|||||
Db 3232 ACAGTGGTCCAGCTCTGGAGCGGTTCTACGACCCCTTGGCAGGGAAGTGTCTGTGAT 3291
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
|||||
Db 3292 GGCAAGAAATAAGACGACTGAATGTCAGTGGCTCCGAGCACACCTGGGCATCGTGCC 3351
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
|||||
Db 3352 CAGGAGCCCATCTCTGTTGACTGCAGCATTCGCTGAGAACATTCGCTATGGAGACAACAGC 3411
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
|||||
Db 3412 CGGTGGTGTGCACGGAAGAGATCGTGAGGCGCAGCAAGGAGGCAACATACATGCTCTC 3471
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
|||||
Db 3472 ATCCAGTCACTGCTCTATAATATATAGCACTAAGTAGAGACAAGAACTCAGCTCTCT 3531
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
|||||
Db 3532 GGTGCCAGAAAACAACGATTCGATTCGATAGCTGCTGCTGCTGCTGATGATATTTTG 3591
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
|||||
Db 3592 CTTTGTGATGAAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGAGGTGTCCAAGAGCC 3651
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 3652 CTGACAAAGCCAGAGAGCCGACCTGCTGATGTTGCTCACCAGCTGTCCACCATC 3711
QY 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||

Db 3712 CAGATCCAGACTAATAGTGGTGTTCAGAAATGGCAGATCAGGAGCATGGCAGCAT 3771
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 3772 CAGCAGCTGCTGGCACAGAAAGCATCTATTTTCAATGTCAGTGTCCAGGCTGGAACA 3831
QY 1280 LysArg 1281
|||||
Db 3832 AAGCGC 3837

RESULT 10
E02326
LOCUS E02326 4378 bp RNA linear PAT 29-SEP-1997
MULTIDRUG resistance relating gene derived from human normal cells.
ACCESSION E02326
VERSION E02326.1 GI:2170561
KEYWORDS JP 1990100680-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ueda,K. and Komano,T.
TITLE HUMAN NORMAL CELL-DERIVED MDR RELATED GENE
JOURNAL Patent: JP 1990100680-A 1 12-APR-1990;
SUNTORY LTD
COMMENT
OS Homo sapiens
PN JP 1990100680-A/1
PD 12-APR-1990
PF 05-OCT-1988 JP 1988251475
PI UEDA KAZUMITSU, KOMANO TORU
PC C12N15/12,C12N1/21,C12Q1/68;
CC strandedness: Single;
CC topology: Linear;
CC *source: tissue=Adrenal gland;
CC *source: clone=SAM1132;
FH Key Location/Qualifiers
FT 5'UTR 1..137
FT CDS 138..3980
FT /gene='Multidrug resistance relating gene' FT
mat_peptide 138..3977
FT /gene='MDR1',
FT /gene='Multidrug resistance relating gene' FT
FEATURES
FT 3'UTR 3981..>4378.
Source
1..4378
Location/Qualifiers
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 1315 a 818 c 1057 g 1188 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4378
Score: 5663.00 Matches: 1163
Percent Similarity: 95.09% Conservative: 56
Best Local Similarity: 90.72% Mismatches: 59
Query Match: 90.58% Indels: 4
DB: 6 Gaps: 4
US-09-672-725C-25 (1-1281) x E02326 (1-4378)
QY 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
|||||
Db 138 ATGGATCTTTGAAGGGGACCGCAATGGAGGAGCAAGAGAACTTTTAACTGAAC 197
QY 20 LysLysSerLysLysGluLysLysGluLysProThrValSerThrPheAlaMet 39
|||||
Db 198 ATAAAGT---GAAAAGATAGAGGAAAGAAACCAACTGTGCTGATTTTCAATG 254
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
|||||
```

Db 255 TTTGCGTATTCAAAATGGCTGCACAAGTTGTATATGGTGGGAACCTTTGGCTGCCATC 314
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 315 ATCCATGGGCGTGGACTTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 374
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 375 GCANATGCAGGA--AATTTAGAACTCTGATGTCACAAACATCACTAATAGAAGTGATATC 431
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyAlaTyrTyr 119
Db 432 AATGATACAGGGTCTTCATGAAT---CTGGAGGAAGACATGACCAAGGTATGCCTATTAT 488
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 489 TACAGTGGAAATTTGGTCTGGGTGCTGGTGTCTTACATTACAGTTTCATTTTGGTGC 548
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPheHisAlaIleMetArg 159
Db 549 CTGGCAGCTGGAAGACAAATACACAAATACAAAACAGTTTTTTCATGCTATATATGCGA 608
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 609 CAGGAGATAGGCTGTTTGTATGTGCACGATGTTGGGGAGCTTAAACACCCGACTTACAGAT 668
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 669 GATGCTCCACAGATTAATGAGGAATTTGGTGACAAAATTTGGAATGTTCTTTCAGTCAATG 728
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 729 GCAACATTTTTCACCTGGTATTATAGTAGGATTTACACGTGGTTGGAAGCTAACCCCTTGTG 788
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 789 ATTTTGGCCATCAGTCCCTCTTCTGGACTGTCTGCTGCTGTGGGCAAGATATCTATCT 848
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 849 TCATTTACTGTATNAGAACTCTTAGCGTATGCCAAAGCTGGAGCAGTAGCTGAAGAGGTC 908
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 909 TTGGCAGCAATTAGAACTGTGATTTGTCATTTGGAGGACAAAGAAAGAACTTGAAGAGTAC 968
Qy 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 969 AACAAAATTTAGAAAGCTAAAGAAATTTGGGATTAAGAAAGCTATTACACCCAATATT 1028
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 1029 TCTATAGGTGCTGCTTCTGCTGATCTATGATCTTATGCTCTGGCCCTCTGGTATGGG 1088
Qy 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 1089 ACCACCTTGGTCTCTCAGGGGAATATTCTATTGGACAAAGTACTCCTGATTTCTTTCT 1148
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1149 GTATTAAATTTGGGCTTTTGTGTGGACAGGCATCTCCCAAGCATTTGAAGCATTTGCAAT 1208
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1209 GCAAGGAGCAGCTTATGAATCTTCAAGATAATTTGATAAAGCAAGTATTGACAGC 1268
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1269 TATTTCGAAGAGTGGGCACAAACAGATAATATTAAAGGGAATTTTGAATTCGAATGTT 1328
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1329 CACTTCAGTTACCCATCTCGAAAAGAAGTTAGATCTTGAAGGCTTGACCTGAAGGTG 1388

Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1389 CAGAGTGGCAGACGGTGGCCCTGGTTGGAAACAGTAGTGGCTGTGGGAAGACACACAGTC 1448
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1449 CAGCTGATGACAGAGCTCTATGACCCACACAGGGGATGTCAGTGTGTATGACAGAGGAT 1508
Qy 460 IleArgThrIleAsnValArgHisLeuArgGluIlePheGlyValValSerGlnGluPro 479
Db 1509 ATTAGGACCAATAATGTAAGCTTCTACGGGAATCATTTGCTGTGTGCTAGTCAGGAACCT 1568
Qy 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1569 GTATTGTTGCCACACAGTAGCTGAAACATTCGCTATGCGCTGAAATGTCAACATG 1628
Qy 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1629 GATGAGATTGAGAAAGCTGTCAAGGAAGCCATGCCCTATGACTTTATCATGAACACTGCC 1688
Qy 520 AsnLysPheAspThrLeuValGlyArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1689 CATAAATTTGACACCCCTGGTTGGACAGAGAGGGCCCGCAGTTGAGTGTGGGCGAGAGCAG 1748
Qy 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 1749 AGGATCGCCATTGACGCTGGCTGTTTCGTAACCCCAAGATCCTCTGCTGGATGAGGCC 1808
Qy 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 1809 ACGTGACGCTTGGACACAGAAAGCGAGCAGTGGTTGAGTGGCTCTGGATAGGCCAGA 1868
Qy 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 1869 AAAGCTCGGACCACTTGTGATAGCTCATCGTTGCTACAGTTCGTAATGCTGACGTC 1928
Qy 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 1929 ATCGCTGTTTCGATGATGGAGTCATTGTCGAGAAGAAATCATGATGAACATCATGAAA 1988
Qy 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 1989 GAGAAAGCATTTACTTCAAACTTGTCAATGTCAGATGCAGACAGCAGGAATGAAGTTGAATTA 2048
Qy 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2049 GAAAATGCAGCTGATGAATCCAAAAGTGAATTCATGCTTGGAAATGCTTCAAAATGAT 2108
Qy 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2109 TCAAGATCCAGCTTAATAAGAAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAGCC 2168
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2169 CAAGACAGAAAGCTTAGTACCAGAGAGCTCTGGATGAAGATATACCTCCAGTTCCCTTT 2228
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2229 TGGAGGATTATGAAGCTAAATTTAACTGAATGGCTTATTTTGTGTTGGTGTATTGTTGT 2288
Qy 720 AlaIleIleAsnGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2289 GCCATTATAATGAGGCGCTCAACACAGCATTTGCAATAATATTTTCAAAGATTATAGGG 2348
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2349 GTTTTACAGAAATGATGATCTCTGAACAAACACAGCAATAGTAACCTGTTTTCCTACTA 2408
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2409 TTGTTTCTAGCCCTTGGAAATTTATTTCTTTTATTACATTTTTCCTTTCAGGGTTTTCACATT 2468

```
QY 780 GlyLysAlaGlyGluLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
|||||
Db 2469 GCGAAGCTGGAGAGATCTCCACCAAGCGGCTCCGATACATGGTTTCCGATCCATGCTC 2528
QY 800 ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
|||||
Db 2529 AGACAGGATGTGAGTTGGTTTGTATGAGCCTAAAAACACCACTGGAGCATTGACTACCAGG 2588
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
|||||
Db 2589 CTCGCCATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTAATATACC 2648
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeu 859
|||||
Db 2649 CAGAAATAGCAATCTGGGACAGGAATAATATATATCCCTCATCTATGGTTGGCAACTA 2708
QY 860 ThrLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLys 879
|||||
Db 2709 ACATGTTTACTCTTAGCAATTGTACCATCATTCGCAATAGCAGGAGTTGTGAAATGAAA 2768
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 899
|||||
Db 2769 ATGTTGCTGCACAGACACTGAAGAATAAGAAAGACTAGAAAGTTCTGGGAAAGATCGCT 2828
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
|||||
Db 2829 ACTGAACAATAGAAAACCTCCGAAACCGTTGTTCTTTGACTCAGGAGCAGAAAGTTTGAA 2888
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
|||||
Db 2889 CATATGTATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATC 2948
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 2949 TTTGGAATTTACATTTTCCCTTCAACCGGCAATGATGATTTTCCATGCTGATGATTTTC 3008
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
|||||
Db 3009 CGGTTTGGAGCCTACTTGTGGTGGCACAATAACTCATGAGCTTTGAGGATGTTCTGTTAGTA 3068
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
|||||
Db 3069 TTTTACGCTGTGTCTTGTGGTGCATGGCGTGGGCAAGTCAAGTTCATTTGCTCCTGAC 3128
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
|||||
Db 3129 TAGTCCAAAGCCAAATATCAGAGCCCATCATATGATCATTTGAAAAACCCCTTTG 3188
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
|||||
Db 3189 ATTGACAGCTACAGCAGGAAGCCCTAATGCCGAACACATTTGAAGAAATGTCACATTT 3248
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
|||||
Db 3249 GGTGAAGTTGATTTCACTATCCACCCGACCGACATCCAGTGTCTCAGGAGCTGAGC 3308
QY 1060 LeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
|||||
Db 3309 CTGAGGTGAAGAGGCCAGACGCTGGCTCTGCTGGCGACGAGTGGCTGTGGGAAGAGC 3368
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
|||||
Db 3369 ACAGTGTCCAGCTCTCGGAGCGGTTCTACAGCCCTTGGCAGGAAAGTGTGCTGTGAT 3428
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
|||||
Db 3429 GCGCAAGAAATAAGCGACTGAATGTTCACTGGCTCCGAGCACACCTGGGCATCGTGCC 3488
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
|||||
Db 3489 CAGAGCCCATCTCTGTTGACTCGACATGCTGAGAACAATTGCCATGGAGACACAGC 3548
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisLysPhe 1159
|||||
Db 3549 CGGGTGGTGCACAGAAGAGATTGTAGGGCAGCAAGAGAGCCCAACATACATGCTTC 3608
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
|||||
Db 3609 ATCGAGTCACTGCTTAATAATATAGCACTAAAGTAGGACAAAGAACTCAGCTCTCT 3668
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
|||||
Db 3669 GGTGGCCAGAAACACGCCATTGCCATAGCTGCTGCCCTTGTTAGACAGCTCATATTG 3728
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
|||||
Db 3729 CTTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGGTTGTCACAGAGCC 3788
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 3789 CTGGCAAAAGCCAGAGAGCGCACCTGCATTTGTGATTGCTCAGCGCTGCCACCATC 3848
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||
Db 3849 CAGAAATGCAGACTTAATAGTGTGTTTTCAGAAATGCAGAGTCAAGAGCATGGCACGCAT 3908
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 3909 CAGCAGCTGCTGGCAGAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGCTGGAACA 3968
RESULT 11
AXI08654
LOCUS AXI08654 4186 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 1 from Patent WO0123565.
ACCESSION AXI08654
VERSION AXI08654.1 GI:13923886
KEYWORDS crab-eating macaque.
SOURCE Macaca fascicularis
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
1 (bases 1 to 4186)
Stockert,P.J., Steimel-Crespi,D.T. and Crespi,C.L.
P-glycoproteins from macaca fascicularis and uses thereof
Patent: WO 0123565-A 1 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
Location/Qualifiers
1..4186
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
100..3942
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37764.1"
/db_xref="GI:13923887"
/translation="MDLEGDRNGAEGKKNFFKLNKSKDKKPKPTVSFVSMFRYSN
WLDKLYVWVGTAAI IHGAGLPLMLVFGMDTDFANAGNLGDLGALLTNSNTDTV
PVMLEDMTRYAVYYSIGAGVLVAAV IOVSFWCLAAAGROIHKIKQFFHAIHQEI
GWFVDVHVGELNTRLTDDSKINGIGDK IGMPFQSMATPTFTGIVGTFGKWLTLVI
LAIISPLGLSAAWAKILVDSFKDELLAYAGAVAEVLAARTVIAFGGKKELER
YKNLEAKRIGIKKAITANISGAFLLIYASVALFWYGTTLVLSKYSIGQVLTV
FFVSLIGAFSVGQSPSIEAFANAGAAFEIKIDNKPSIDSYSKSGHPRDKNL
FRNWHFSYPSRKEVKILGKLNKVSQGTVALVNSGCGKSTTVOLMORLYDPTGCM
VSVQGDQIRINVRFLREILGVVSOEPLFATRIENIRYREDVTMDTEKAVKAN
AYDITMKLPQKFDTLVGERGAQLSGGOKRIATARALVRNPKILLDEASALDTSE
AVQVADKARKKRTTIVIAHRUSTVRNADVIAGDDGVIVKGNHDELMKRGIVFK
LVTMQATNEIELENAASESETIDTMSHDGSSGLIRKSTRSRVSQGDRLK
STKEALDESIPPSVFRWIKMLNLTWEPYVGVFCALINGLQPAFAVPSKIGFT
RNDIAETKRONSLFSLFLVLGVTFITFLOGTFTGKAGEILTTLRLRYWFRSMRLR
QDVSEFDDPKNTTCALTRLANDAAOVKGAIGSLAITONIANIGTGIILYIGHO
LTLULLAIVPIIATAGVEMKMLSGQALKKKKELEGAKTATEAIENTFRVTVSLTQEQ
```

KPEHMYDQSLQVPYRNSLRKAHIFGCTIFSTQAMMYFSYAGCFRFGALYLVHLSMSPF
DVLIVFSAVFGAMAGVQSSFPDYAKAKVSAHHIIMITEKTLIDSYSTEGGLKPNP
LEGWTFNEVFNYPTRLDIPVLOGLSLEVKKGOTLALVSGSGGKSTVVQOLLERFYD
PLAGKVLDDGEIKQLNQMWRRAHLGIUSQEPILFDCSISENIAYGNSRVSQSEELV
RAAKEANTHAFIESLPNKYSITRVGDKGTQLSGGOKRIATARALVRPHILLUDEATS
ALDTESEKRVQEALDKAREGRTIVIAHRLSTIQIONADLIVVFQNGRVKEHGHQOLLA
OKGIYFMSVQVQAKRQ"

BASE COUNT 1226 a 799 c 1039 g 1122 t
ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4186
Score: 5862.00 Matches: 1163
Percent Similarity: 95.32% Conservatives: 59
Best Local Similarity: 90.72% Mismatches: 56
Query Match: 90.56% Indels: 4
DB: 6 Gaps: 4

US-09-672-725C-25 (1-1281) x AX108654 (1-4186)

QY 1 MetAspProGluGlyArgLysGlySerAlaGlu---LysAsnPhetripLysMetGly 19
Db 100 ATGATCTTGAAGGGACCGCAATGAGGAGCAGAGAGAAAGAACTTTTAACTGAC 159
QY 20 LysLysSerLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 160 AATAAAGT---AAAAAGATAAGAGAGAAAGAAACCAACTGTCAGTGTATTTTCAATG 216
QY 40 PheArgTyrSerAsnTrpLeuaspargLeuTyrMetLeuValGlyThrMetAlaIle 59
Db 217 TTTTCGTATTCAAATTTGGCTTGACAAAGTTGTATATGTTGGTGGAACTTTGGCTGCCATC 276
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 277 ATCCATGGAGCTGGACTTCTCTCATGATGCTGGTGTGGAGACATGACGGATACCTTT 336
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 337 GCAAAATGCAGGA---AATTTAGGAGATTTAGGAGCTCTGTTGACTAATAGCAGTAATATC 393
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 394 ACTGATACAGTCCCGTCATGAAT---CTGGAGGAAGATATGACCAAGGTATGCCATATAT 450
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 451 TACAGTGGAAATTCGTGCTGGGGTCTGGTGTGCTGCTTACATCAGGTTTCATTTGGTGC 510
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 511 CTGGCAGCTGGAAGACAAATACACAAATATAGAAAACAGTTTTTTCATCTATAATGCCA 570
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyLeuAsnThrArgLeuThrAsp 179
Db 571 CAGGAGATAGCGTGGTGTGATGTGCAGATGTTGGGGAGCTTAACACCGCGCTTACAGAT 630
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db 631 GATGCTCCAAGATTAAATGAAGCAATTTGTGACAAAATTTGGAATGTTCTTTTCAGTCAATG 690
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 691 GCAACATTTTTCATCTGGGTTTATAGTAGGATTTACACGTTGGTGGAAAGTAAACCCCTTGG 750
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 751 ATTTTGGCCATCAGTCTCTGTTCTTGACCTGCTCAGCTGCGACTGCGGCAAGATAGTCT 810
QY 240 SerPheThrAspLysGluLeuValaTyrAlaLysAlaGlyAlaValaGluGluVal 259
Db 811 TCATTTACTATAAGAACTCTTAGCTTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 870
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyr 279

Db 871 TTGCAGCAATTAGAACTGTGATTTGGAGGACAAAAGAAAGAACTCGAAAGGTAC 930
QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle 299
Db 931 AACAAAATTTAGAAAGAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCAATATT 990
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 991 TCTATAGTGTGCTTTTCTCTTATCTATGTCATCTTATGCTCTGGCTTCTGTATGGG 1050
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 1051 ACCACCTTGGTCTCTCAAGGAATATTTATTGGACAAGTACTACTGTATTCTTTCT 1110
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1111 GTATTAAATGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCCAAT 1170
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1171 GCAAGAGGAGCAGCTTTTGAATCTTCAAGATAATTGATAATAAGCCAGTATTGACAGC 1230
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1231 TATTGGAAGAGTGGGCACAAACCAAGATAAATTAAAGGAAATTTGGAATTCAGAAATGTT 1290
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1291 CACTTCAGTATACCATCTCGAAAAGAGATTAAAGATCTTGAAGGCGCTGAACCTGAAGGTG 1350
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1351 CAGAGTGGCAGACGGTGGCCCTGTTGGAAACAGCGCTGTGGGAAAGACACACCGGTC 1410
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1411 CAGCTGATGACAGAGCGCTTTATGACCCACAGAGGCGATGGTCACTGTTGATGGACAGAT 1470
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1471 ATTAGGACCAATAACGTAAGGTTTCTACGGGAAATCATCGGTGTGTGAGTCAGGAACCT 1530
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1531 GTATTGTTGGCCACACGATAGCTGAAACATTCGCTATGCTGCTGGAAGATGTCACCATG 1590
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1591 GATGAGATTGAGAAGCTGTCAGGAAGCCAAATGCCCTATGACTTTATCATGAAGTGCCT 1650
QY 520 AsnLysPheAspThrLeuValGlyArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1651 CAGAAAATTTGACACCTCTGTTGGAGAGAGAGGGGCCAGCTGAGTGGTGGGACAGAGCAG 1710
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAla 559
Db 1711 AGGATCGCAATGACAGTGGCCCTGTTGCGAACCCCAAGATCCTCCTGCTGGAGAGGCC 1770
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 1771 ACGTCAGCTTGGACACAGAAAGTGAAGCAGTGGTTCAGGTTGGCTCTGGATAAGGCCAGA 1830
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 1831 AAAGTCGGACCAACCATTTGATAGCTCATCGTTTGTCTACGGTTCGTTAATGCCGAGTC 1890
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 1891 ATCGCTGTTTTCGATGATGGAGTCATTGCGAGAAAGGAAATCATGATGACCTCATGAAA 1950
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639

Db 1951 GAGAAAGCATTACTTCAAACTTGTCAATGTCAGACAGCAGGAAATCAAAATTGAATTA 2010
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2011 GAAATGACGTGATGAATCCAAAAGTGAATGTATACCTTGGAAATGCTTCACATGAT 2070
QY 660 SerGlySerSerLeuLeuLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2071 TCAGGATCCAGTCTAATAAGAAAAGATCCACTCGTAGSAGTGTCCGTGGATCACAAGC 2130
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGlnValProProValSerPhe 699
Db 2131 CAAGACAAAGAGCTTAGTACCAAGAGAGCTCTGGATGAAGATATACCTCCAGTTTCCCTT 2190
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyLysPheCys 719
Db 2191 TGGAGGATTAAGCACTAAATTAAGTACTGAGTGGCCCTTAATTTGTTGGTGAATTTGT 2250
QY 720 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2251 GCCATTATAAATGGAGGTCTGCAACACAGCATTTGGCAGTAATATTTTCAAGATATATAGG 2310
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2311 ATTTTACAAAGAAATGATGATGCCGGAACAAAACACAGAAATAGTAATCTGTTTTCATA 2370
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2371 TTGTTTCTAGCTTGGAAATGTTTCTTTTATATACATTTTTCCTTCAGGGCTTCACATTT 2430
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2431 GGCAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATGGTTTCCGATCCATGCTC 2490
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2491 AGACAGGATGTGAGCTGTTGTGATGACCTTAAAGCAACACCACTGGAGCATTTGACTACCAGG 2550
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2551 CTCGCCATGATGCTGCTCAAGTTAAAGGGCTTAGTTCAGGCTTGTCTATTAATATACC 2610
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2611 CAGAAATAGCAAAATCTTGGACAGGAATAATATATCTTAATCTATGTTGGCACTG 2670
QY 860 ThrLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLys 879
Db 2671 ACACCTGTTACTCTTAGCAATTTGTACCCATCATTTGCAATPAGCAGGAGTTGTTGAAATGAA 2730
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2731 ATGTTGCTGGACAGCACTGAAGATTAAGAAAGAACTAGAAAGTGTCTGGGAATCGCT 2790
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2791 ACTGAAGCAATAGAAAACCTCCGAACCTGTTCTTCTTGACTCAGGACGACAAAGTTGAA 2850
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2851 CATATGATGATCAGAGTTTCAGGTACCATACACAGAAACTCTTTGAGGAAAGCACATC 2910
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2911 TTTGGAATCAGCTTTTCTCCACGAGCAATGATGATATTTTCTATGCTGGAGTGTTC 2970
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 2971 CGGTTGGAGCCTACTTGGTGGCACAATAGTCTCATGAGCTTTGAGGATGTTCTCTTAGTA 3030
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3031 TTTTCAGCTGTTGTTGTTGGTGGCATGGCCGTGGGCAAGTCAGTTTCATTTGCTTCCCTGAC 3090

QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3091 TATCCAAAGCAAGATATCAGCAGCCACATCATGATCATTTAAAAACCCCTTTG 3150
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3151 ATTGACAGCTACAGCAGACAGAGGCCTAAAGCCGAACACATTGGAAGGAATGTACATTT 3210
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3211 AATGAAGTTGTATTCAACTATCCACCCGAGCTGGACATCCAGTGTCTTCAAGGCGTGAGC 3270
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3271 CTGGAAGTGAAGAGGCCAGACGCTGGCCCTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3330
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3331 ACGTGTGTCCAGCTCTCTGGAGCGGTTCTATGACCCCTTGGCGGGAAGTGTCTGTGAC 3390
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3391 GGCAAGAAATAAAGCAACTGAATGTTCAGTGGCTCCGAGCACACCTGGGCATCGTGTC 3450
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3451 CAGAGCCCATCTCTGTTGACTGCAGCATTAGTGAGAACATTCCTATGGAGACACAGC 3510
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3511 CGGTGTGTGCACAGGAAGAGATCGTGAGGCGCAGCCAAAGGAGGCAATATACAGCCTTC 3570
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3571 ATCGAGTCACTGCTATAAATATAGCACCAGAGTAGGAGACAAAGAACTCAAGCTCTCT 3630
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3631 GGTGGCCAGAAACACGCATTCCTAGCTAGCTCGTGGCCCTTTGTAGCAGCCTCATATTTG 3690
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3691 CTTTGTGATGAAGCCACATCAGCTCTGGATACAGAAAGTGAAGGTTGTCCAAGAGCC 3750
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3751 CTGGACAAAGCCAGAGAGCGCTACCTGCATTTGTGATTGCTCAGCGCTGTCCACCATC 3810
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3811 CAGAATCAGACTTAATAGTGGTGTTCAGATGGCAGAGTCAAGGAGCAGCGCACACAT 3870
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3871 CAGCAGCTGTGGCAGACAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAGCA 3930
QY 1280 LysArg 1281
Db 3931 AAGCGC 3936

RESULT 12

AX108656
LOCUS AX108656 4195 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 3 from Patent WO0123565.
ACCESSION AX108656
VERSION AX108656.1 GI:13923888
KEYWORDS crab-eating macaque.
SOURCE Macaca fascicularis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Macaca.


```

REFERENCE 1 (bases 1 to 4195)
AUTHORS Stocker,P.J., Steimel-Crespi,D.T. and Crespi,C.L.
TITLE P-glycoproteins from macaca fascicularis and uses thereof
JOURNAL Patent: WO 01233565-A 3 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
    source      Location/Qualifiers
    1..4195     Location/Qualifiers
                /organism="Macaca fascicularis"
                /db_xref="taxon:9541"
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAC37765.1"
                /db_xref="GI:13923889"
                /translation="MDLEGDRNGAEKKNFFKLNKKKKKKKPKTPVSFSPMPSYN
                WLDKLYMVVGLAIIHGAGLPLMWLVEGDMTDIFANAGNLGDLGALLFNNTNSNIT
                DTVPVNMLEDMTRIATYISGIGAGLVVAAYIQVSFCLAGRLHKKRQFFHAINR
                QIVFDVHDVDELNTRLDVDSKINEGIDKIGMFFQSMATFFTFIVGTRGKWL
                LVLAISPLGLSAVWAKILSSFTDKELLAYAKAGAAVEVLAARIVTAFGQKKE
                LERYNKNLEAKRIGIKKAITANISIGAEFLLIYASALAFWYGTTLVLKEYSIGOV
                LTVFSLVIGAFSGVQASPTAFANARGAAFEIKIIDNKPISIDSYSKSGHKPDNIK
                GNLEFRNVHFSYSPKKEVKILKGLNKLKVSQGVVALVGNSSCGKSTVQLMORLYDPT
                EGMVSYDQCDIRITINREFRLIIGVVSQEPVLEATTAAENIRYREDVTDWDELEKAV
                ENAYVDFIMKLPQRDTLVBGERGQSLGQKQRTAIALALVRNPKILLLDPAISALDT
                ESEAVVQVALDKARKGRRTTIVIAHRLSTVRNADVIAGFDGVIIVEKGNHDELMKEGI
                YFKLVTMOTAGNEILENAADESKSEIDTLEMSHSDSSLIKRSTRSRVRSQGGD
                RKLSTKEALDESIPVSWRLMKNLNTEWPFVGVFCALINGGLGPAFAVIFSKIIIG
                ITRNDDEATKRONSLFSLFLVLGIVSFITFEFLOGTFGKAGEILTFLKRLRYWFRS
                MLRDVSWFDPKNTGALTTRLANDAAQVKAIGSRLAIITONTANLGNGLIISLY
                GMQLTLLLAIVPIITAIAGVEMKMLSGQALDKKKELEGAKIATEALIENTPRTVSLT
                OQKFEHMYDSQSLVYNSLRKAHIFGITTFQAMMYFSYAGCFRFGAYLVAHSLM
                SPEDVLLVFSVAVFGAMAVGVSSFPADYAKAKSAAHIMIEIKPLIDSYSYEGKL
                PNTLEGNVTFNEVNPYTRLDIPVLQGLSLVKKGOTLALVSGSGCKSTVQOLLER
                FVDPLAGVLLDGKEIKOLNQLWLAHGLIVSOEPLIFDCSISENIAVGDNSRVVSOE
                ELVRAAKEANTHAFTESLPNKYSIFRVGDKTQLSGGQKQRIARALRVOPHILLIDE
                EIALDITSEKVEQALDKAREGRTCVIAHRLSTIQNALDILIVFQNGRVAEHGTHQO
                LLAQKGIYFSVMYSVQAQAKRQ"
BASE COUNT 1230 a 801 c 1039 g 1125 t
ORIGIN

Alignment Scores:
Pred. No.:      0      Length:      4195
Score:          5859.50      Matches:      1165
Percent Similarity: 94.80%      Conservative: 57
Best Local Similarity: 90.38%      Mismatches: 52
Query Match:      90.52%      Indels: 15
DB:              6      Gaps: 5

US-09-672-725C-25 (1-1281) x AX108656 (1-4195)
QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGlu---LysAsnPheTrpLysMetGly 19
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 ATGCATCTTGAGGGGACCGCATGAGGAGGAGCAGAGAAAGAACTTTTAAACGTGAAC 159
QY 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 160 AATAAAGT---AAAAAAGATAAGAGAAAGAAACCAACTGCTCAGTGTATTTTCAATG 216
QY 40 PheArgTyrSerAsnTrpLeuPheArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 217 TTTTCGCTATTCAAAATGGCTTGACAGATTGATATGCTGGTGGGAACCTTTGGCTGCCATC 276
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 277 ATCCATGGACCTGACTTCTCTCATGATGCTGTGCTTTGGAGACATGACGGATACCTTT 336
QY 80 AlaAsnAlaGly-----IleSerArgAsnLysThrPheProVal 92
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 337 GCAAAATGCAGGAATTTAGGAGATTTAGGAGCTCTCTGTTGTTAAACACACACT----- 387
QY 93 IleIleAsnGluSerIleThrAsnAsnThrGlnHisPheIleAsnHisLeuGluGlu 112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 -----AATAGCAGTAATATCACTGTATACAGTCCCGCTCATGAAT---CTGGAGGAAGAT 438

```

```

QY 113 MetThrThrTyrAlaTyrTyrTyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyr 132
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439 ATGACCAAGGTATGCTATTATTACAGTGAATGGTCTGGGTGCTGGTGGTCTTAC 498
QY 133 IleGlnValSerPheTrpCysLeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGln 152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 499 ATTCAAGTTTCATTTTGGTGGCTGGCAGCTGGAAGACAAATACACAAAATAGAAAACAG 558
QY 153 PhePheHisAlaIleMetArgGlnGluIleGlyTrpPheAspValHisAspValGlyGlu 172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 559 TTTTTCATGCTATAATGCCACAGCAGATAGGCTGGTTTGTATGTGCACGATGTTGGGAG 618
QY 173 LeuAsnThrArgLeuThrAspValSerLysIleAsnGluGlyIleGlyAspLysIle 192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 619 CTTAACACCCGGCTTACAGATGATGCTCCAAGATTAAATGAAGGAATTTGTCACAAAATT 678
QY 193 GlyMetPhePheGlnSerIleAlaThrPhePheThrGlyPheIleValGlyPheThrArg 212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 679 GGAATGTTCTTTTCAGTCAATGGCAACATTTTCTACTGGGTTTATAGTAGGATTTTACACGT 738
QY 213 GlyTrpLysLeuThrLeuValIleLeuAlaIleSerProValLeuGlyLeuSerAlaAla 232
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 739 GGTGGGAAGCTAACCTTGTGATTTTGGCCATCAGTCTCTTCTTGGACTGTCAGCTGCA 798
QY 233 IleTrpAlaLysIleLeuSerSerPheThrAspLysGluLeuLeuAlaTyrAlaLysAla 252
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 799 GTCTGGGCAAGATACCTCTCTTCACTAATGAAGAACTCTTAGCTTATGCAAAAAGCT 858
QY 253 GlyAlaValAlaGluGluValLeuAlaAlaIleArgThrValIleAlaPheGlyGlyGln 272
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 859 GGACAGTAGCTGAAGAGGCTCTGGCAGCAATTAAGAACTGTGATGCTTTTGGAGGACAA 918
QY 273 LysLysGluLeuGluArgTyrAsnLysAsnLeuGluAlaLysGlyIleGlyIleLys 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 919 AAGAAAGAACTCGAAGGTACACAAAATTTAGAAGCACTAAAGAAATTTGGATAAG 978
QY 293 LysAlaIleThrAlaAsnIleSerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyr 312
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 979 AAAGCTATTACAGCAATATTCTTATAGGTGCTCTTCTTCTTCTTATCTATGCACTTAT 1038
QY 313 AlaLeuAlaPheTrpTyrGlyThrSerLeuValLeuSerSerGluTyrSerIleGlyGln 332
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1039 GCTCTGGGCTTCTGGTATGGGACCACTTGTCTCTCAAGGAATATTTATTTGGACAA 1098
QY 333 ValLeuThrValPhePheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerPro 352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1099 GTACTCAGTGTATCTTTTCTGTATTAATTTGGGCTTTTAGTGTGGACAGGCACTCCA 1158
QY 353 SerIleGluAlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheLysIleAsp 372
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1159 AGCATTTGAAGCATTTGCAAAATGCAAGAGGAGCAGCTTTTGAATCTTCAAGATTAATTGAT 1218
QY 373 AsnLysProSerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGly 392
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1219 AATAGCCCAAGTATTGACAGCTATTCGAGAGTGGGCACAAACCCAGATAATATTAAGGA 1278
QY 393 AsnLeuGluPheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeu 412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1279 AATTTGGAATTCAGAAATGTTCACTTTCAGTTACCATCTCGAAAAAGAAAGTTAAGATCTTG 1338
QY 413 LysGlyLeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGly 432
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1339 AAGGCCCTTGAACCTTGAGGTGAGAGTGGCAGAGCTGGTGGTGGTGGAAACAGCGGC 1398
QY 433 CysGlyLysSerThrThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMet 452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1399 TGTGGGAGAGACCAACCGTCCAGCTGATGAGAGGCTTTATGACCCACAGAGGGCATG 1458
QY 453 ValCysIleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThr 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1459 GTCAGTGTGATGGACAGAGATATTAGGACCATAAACGTAAGGTTTCTACGGGAATCATC 1518

```


QY 473 GlyValValSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyr 492
Db 1519 GGTGTGGTGTAGTCAGGAACCTGTATGTTTCCACACGATAGCTGAAACATTCGCTAT 1578
QY 493 GlyArgGluAsnValThrMetAspGluIleGluLysAlaValLysGluAlaAsnAlaTyr 512
Db 1579 GGTCTGAAGATGTCACCATGGATGAGATTGAGAAAGCTGTCAAGGAAGCCAAATGSCCTAT 1638
QY 513 AspPheIleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGln 532
Db 1639 GACTTTATCATGAACGCCTCAGAAATTTGACACCCTGGTTGGAGAGAGGGGCCAG 1698
QY 533 LeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLys 552
Db 1699 CTGAGTGTGGCAGAGACAGAGGATGCCATTGCACGTGCCCTGGTTGCGCAACCCCAAG 1758
QY 553 IleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGln 572
Db 1759 ATCTCTCTGTGGACGAGGCCACGCTCAGAAATTTGACACCCTGGTTGGAGAGAGGGGCCAG 1818
QY 573 ValAlaLeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSer 592
Db 1819 GTGGCTCTGGATAGGCCAGAAAGGTCGGACCACTATGATAGCTCATFCGTTGTCT 1878
QY 593 ThrValArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGluLysGly 612
Db 1879 ACGGTTCTGAATCCGACGTCATCGCTGGTTTCGATGATGGAGTCATTGTGGAGAAAGGA 1938
QY 613 AsnHisaspGluLeuMetLysGluLysGlyIleIleIleIleValThrMetGlnThr 632
Db 1939 AATCATGATGAGCTCATGAAGAGAAAGGCAATTTACTTCAAACTGTGCACAAATGCAGACA 1998
QY 633 ArgGluAsnGluIleGluLeuGluAsnAlaThrGlyGluSerLysSerGluSerAspAla 652
Db 1999 GCAGAAATGAATGTAGTAAGAAATGCAGCTGATGAATCCAAAGATGAATGTATACC 2058
QY 653 LeuGluMetSerProLysAspSerGlySerLeuIleLysArgArgSerThrArgArg 672
Db 2059 TTGGAATGTCTTACATGATTCAGGATCCAGTCCAGTCTAATAAGAAAGATCCACTCGTAGG 2118
QY 673 SerIleHisAlaProGlnGlnAspArgLysLeuGlyThrLysGluAspLeuAsnGlu 692
Db 2119 AGTCTCGGTGGATCACAAGGCCAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAA 2178
QY 693 AsnValProValSerPheTrpArgIleLeuLysLeuAsnSerThrGluTtpProTyr 712
Db 2179 AGTATACCTCCAGTTCTCTTTGGAGGATTATGAAGCTAAATTAACGTAGTGCCCTTAT 2238
QY 713 PheValValGlyIlePheCysAlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIle 732
Db 2239 TTTCTGTGTGTATTTTGGCCATTATAAATGGAGGCTCTGCAACCCAGCATTTGCAGTA 2298
QY 733 IlePheSerArgIleIleGlyIlePheThrArgAspGluAspProGluThrLysArgGln 752
Db 2299 ATATTTTCAAAGATTATAGGATTTTACAAAGAAATGATGATGCCGAACAAACACGACAG 2358
QY 753 AsnSerAsnMetPheSerValLeuPheLeuValIleuGlyIleIleSerPheIleThrPhe 772
Db 2359 ANTAGTACTGTGTTTTCACATGTTTCTAGTCTTGGAAATGTGTTCTTTTATACATTT 2418
QY 773 PheLeuGlnGlyPheThrPheGlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyr 792
Db 2419 TTCTCTCAGGGCTTCACATTTGGCAAGCTGGAGAGATCCTCACCAGCGGCTCCGNATC 2478
QY 793 MetValPheArgSerMetLeuArgGlnAspValSerTrpPheAspAspProLysAsnThr 812
Db 2479 ATGGTTTTCGGATCCATGCTCAGACAGATGTGAGCTGGTTGTATGACCCCTAAACACACC 2538
QY 813 ThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaGlnValLysGlyAlaIleGly 832
Db 2539 ACTGGAGCATTTGACTACGAGCTGCGCAATGATGCTCTCAAGTTAAAGGGGCTATAGGT 2598
QY 833 SerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSer 852

Db 2599 TCCAGGGTTCCTATAATATCCAGAAATATAGCAAAATCTTGGGACAGAAATAATATATACC 2658
QY 853 LeuIleIleGlyTrpGlnLeuThrLeuLeuLeuAlaIleValProIleIleAlaIle 872
Db 2659 TTAATCTATGGTTGGCAACTGACACTGTACTCTTAGCAATGTACCCATCATTTGCAATA 2718
QY 873 AlaGlyValValGluMetLysMetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeu 892
Db 2719 GCAGGAGTTGTTGAAATGAAATGTTGCTGGACAGCACTGAAAGATTAAGAAAGAACTA 2778
QY 893 GluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeu 912
Db 2779 GAAGGTGCTGGGAAGATCGCTACTGAAGCAATAGAAACTTCCGAACCTGTTGTTCTTTG 2838
QY 913 ThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsn 932
Db 2839 ACTCAGGACGCAAGATTGAACATATGTATGATCAGAGTTTCAGGTACCATACAGAAAC 2898
QY 933 SerLeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyr 952
Db 2899 TCCTTTCAGGAACACACATCTTTGGAATCACGTTTTCCTTCACGAGCAATGATGAT 2958
QY 953 PheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsn 972
Db 2959 TTTTCTCTATGCTGATGTTTCCGGTTTGGAGCTACTTGTGGCACATAGTCTCATGAGC 3018
QY 973 PheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGln 992
Db 3019 TTTGAGGATGTTCTGTGTAGTATTTTCAGCTGTGTTGTTGGTCCCATGCCGTGGGCAA 3078
QY 993 ValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMet 1012
Db 3079 GTCAGTTCAITTCCTCCTGACTATGCCAAAGCCAAAGATATCAGACCCACATCATCATG 3138
QY 1013 IleIleGluLysProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThr 1032
Db 3139 ATCATTTGAAAAACCCCTTTGATTGACGCTACAGCACAGAAAGGCTTAAGCCGAACACA 3198
QY 1033 LeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIle 1052
Db 3199 TTGGAAGGAAATGTCATTTAATGAAGTTGATTCAACTATCCCCACCCGATGGCAATC 3258
QY 1053 ProValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrLeuAlaLeuValGly 1072
Db 3259 CCAGTCTTCAGGGGCTGAGCCTGGAAAGTGAAGAGGGCCACAGCTGCCCTGGTGGC 3318
QY 1073 SerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeu 1092
Db 3319 AGCAGTGGCTGTGGGAAGACGACGGTGGTCCAGCTCCTGGAGCGGTTCTATGACCCCTG 3378
QY 1093 AlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArg 1112
Db 3379 GCGGGAAAGTGTGTTGACGCAAGAAATAAAGCAACTGAATGTTTCAGTGGTCCGA 3438
QY 1113 AlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsn 1132
Db 3439 GCACACTGGGATCTGTGCCAGGAGCCCATCTCTGTTTGTAGCTGCAGCATTAGTAGAAC 3498
QY 1133 IleAlaTyrGlyAspAsnSerArgValValSerHisGluGluIleMetGlnAlaLys 1152
Db 3499 ATTGCTCATGGAGACAACACCGGGTGGTGTCCACAGGAAGACATCGTAGGCGCACCAAG 3558
QY 1153 GluAlaAsnIleHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGly 1172
Db 3559 GAGGCCAATATACACGCTTCATCGAGTCACGTGCCTTAATAAATATAGCACAGAGTAGGA 3618
QY 1173 AspLysGlyThrGlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeu 1192
Db 3619 GACAAAGGAACTCAGCTCTCTGGTGCCAGAAACCAACGCAATGGCCATAGCTCGTGCCCTT 3678
QY 1193 ValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer 1212

Db 3679 GTTAGACAGCCTCATATTTTGGTGAAGCCACATCAGCTCTGGATACAGAAAGT 3738
Qy 1213 GluLysValValGlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIle 1232
Db 3739 GAAAGGTTGTCCAGAGACCCCTGCACAAAGCCAGAGAGGCCGTACCTGATGTGATT 3798
Qy 1233 AlaHisArgLeuSerThrIleGlnAsnAlaAspLeuIleValValPheGlnAsnGlyLys 1252
Db 3799 GCTCACCGCCTCTCCACCATCAGAAATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGA 3858
Qy 1253 ValLysGluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMet 1272
Db 3859 GTCAAGGACACGGCACATCAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATG 3918
Qy 1273 ValSerValGlnAlaGlyAlaLysArg 1281
Db 3919 GTCAGTGTCCAGCGTGGAGCAAGCGC 3945
RESULT 13
AX322789
LOCUS AX322789 3860 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 3 from Patent WO0192877.
ACCESSION AX322789
VERSION AX322789.1 GI:18093767
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE Sorrentino, B. and Schuetz, J.
AUTHORS Method of identifying and/or isolating stem cells
TITLE Patent: WO 0192877-A 3 06-DEC-2001;
JOURNAL ST. JUDE CHILDREN'S RESEARCH HOSPITAL (US)
FEATURES Location/Qualifiers
source
1..3860
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1135 a 746 c 956 g 1023 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3860
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 6 Gaps: 4
US-09-672-725c-25 (1-1281) x AX322789 (1-3860)
Qy 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1 ATGGATCTTGAAGGGACCGCAATGGAGGAGCAAGAGAGAACTTTTAAACTGAAC 60
Qy 20 LysLysSerLysLysGluLysGluLysLysGluLysProThrValSerThrPheAlaMet 39
Db 61 AATAAAAGT---GAAAAAGATAAGAGGAAAAAGAAACCAACTGTCAGTGTATTTTCAATG 117
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 118 TTTTCGCTATTCAAAATGGCTTACCAAGTTGTATATGGTGGGAACCTTTGGCTGCCATC 177
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 178 ATCCATGGGGCTGGACTTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 237
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 238 GCAAAATGCAGGA---AATTTTAAAGATCTGTGTCACAAACATCACAATAGAGGTATATC 294
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 119

Db 295 AATCATACAGGGTCTTCATGAAT---CTGGAGGAAGACATGACCAGATATGCTTATTAT 351
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 352 TACAGTGAATTTGGTGTGGGTGCTGGTGTGCTTACATTACAGTTTCATTTTGGTGC 411
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 412 CTGSCAGCTGGAAGACAAATACACAAAATPAGAAAACAGTTTTTTTCATGCTATAATGCGA 471
Qy 160 GlnGlnIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 472 CAGGAGATAGCGTGGTTGATGTGCAGATGTTGGGAGCTTAACACCCGACTTACAGAT 531
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 532 GATGCTCTTAAGATTAAATGAAGTTATTGGTGACAAAATTTGGAATGTTCTTTTCAGTCAATG 591
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 592 GCAACATTTTTCAGTGGGTTTATAGTAGGATTTACACGTGTTTGGAGCTTAACCCCTGTG 651
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 652 ATTTTGGCCATCAGTCTCTGTTCTTGGACTGTGACTGCTCTGCGGCAAGATACTATCT 711
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
Db 712 TCATTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTGC 771
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 772 TTGCGACCAATTAGACTGTGATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 831
Qy 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 832 AACAAAATTTAGAGAGCTTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCAATATT 891
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 892 TCTATAGTGCTGCTTTCCTGATCATCTTATGCTCTGGCCTTCTGTGATGGG 951
Qy 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPheSer 339
Db 952 ACCACTTGGTCTCTCAGGGGAATATTTATTGGACAAGTACTACTGATTCTTTCT 1011
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1012 GTATTAAATGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCAAT 1071
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1072 GCAAGAGGAGCAGCTTATGAAATCTTCAAGATATTTGATTAATAGCCAACTATTGACAG 1131
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1132 TATTCSAAGAGTGGGCACAAACCAAGATAATTAAGGAAATTTGGAATTCAGAAATGTT 1191
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1192 CACTTCAGTACCCTCTCGAAAGAAAGTTAAGATCTTGAAGGGCCTGAACCTGAAGGTG 1251
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrVal 439
Db 1252 CAGAGTGGCAGACGGTGGCCCTGGTTGGAACAGTAGTGGCTGCTGGGAAGACACACAGTC 1311
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1312 CAGCTGATGACAGAGCTCTATGACCCACAGAGGGGATGCTCAGTGTGTGATGACAGGAT 1371
Qy 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1372 ATTAGGACCATAAATGTAAGTTTCTACGGGAAATCATTTGCTGTGGTCACTCAGGAACCT 1431


```
Db 3592 CTTTGGATGAAGCCACGCTGACCTGGATACAGAAAGTGAAAGGTTGTCCAGAAAGCC 3651
|||||
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 3652 CTGGCAAAAGCCAGAGAGCCGACCTGCGATGTGATGCTCACCGCTGTCCACCATC 3711
|||||
Qy 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||
Db 3712 CAGAATGCAGACTTAATAGTGTGTTTTCAGAAATGCAGAGTCAAGGACGATGCGCAGCAT 3771
|||||
Qy 1260 GlnGlnLeuAlaGlnLysGlyIleThrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 3772 CAGCAGCTGCTGGCAGAAAGGATCATATTTTCAATGGTCAGTGTCCAGGCTGGAAACA 3831
|||||
Qy 1280 LysArg 1281
|||||
Db 3832 AAGCGC 3837
|||||
RESULT 14
AX336420
LOCUS AX336420 4646 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 6929 from Patent WO0194629.
ACCESSION AX336420
VERSION AX336420.1 GI:18127139
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrihan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 6929 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Location/Qualifiers
Source 1..4646
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1371 a 892 c 1129 g 1254 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4646
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 6 Gaps: 4
US-09-672-725C-25 (1-1281) x AX336420 (1-4646)
Qy 1 MetAspProGluGlyClyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
|||||
Db 425 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAAAGAACTTTTAAACTGAAC 484
|||||
Qy 20 LysLysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
|||||
Db 485 AATAAAAGT---GAAAAGATAGAGAGGAAAGAAACCAACTGTCAGTGTATTTCAATG 541
|||||
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
|||||
Db 542 TTTTCGCTATTCAAAATGGCTTGACAAGTTGTATATGGTGGTGGGAACTTTGGCTGCCATC 601
|||||
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
|||||
Db 602 ATCCATGGGGCTGGACTTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 661
|||||
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
|||||
Db 662 GCAAATGCAGGA---AATTTAGNAGATCTGATGTCAACATCATCAANTAGAGATGATC 718
|||||
```

```
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrThrThrThrThrThr 119
|||||
Db 719 AATGATACAGGGTCTTTCATCAAT--CTGAGAGCAAGCATGACAGGATGCTGCTTAT 775
|||||
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaIleGlnValSerPheTrpCys 139
|||||
Db 776 TACAGTGGAAATTTGGTGTGGGGTGTCTGCTGTCTACATTCAGGTTTCATTTTGGTGC 835
|||||
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
|||||
Db 836 CTGCGACCTGGAAGACAAATACACAAATTAGAAAAACAGTTTTTTCATGCTATAATGCGA 895
|||||
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
|||||
Db 896 CAGGAGATAGGCTGTTTGTATGTGCACGATGTTGGGAGCTTAACACCCGACTTACAGAT 955
|||||
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
|||||
Db 956 GATGCTCTAAGATTAAATGAAGTTATTGGTGACAAAATTTGGAATGTTCTTTCAAGTCAATG 1015
|||||
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
|||||
Db 1016 GCAACATTTTCACTGGTGTATAGTAGGATTTACACGTGCTGGTAGCTAACCCTTGTG 1075
|||||
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
|||||
Db 1076 ATTTTGGCCATCAGCTCTCTTGGACTGCTCAGCTGCTCTGGGCAAGATACTATCT 1135
|||||
Qy 240 SerPheThrAspLysGluLeuAlaIleValGlyAlaGlyAlaValAlaGluGluVal 259
|||||
Db 1136 TCATTTACTATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTTC 1195
|||||
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
|||||
Db 1196 TTGCGACCAATTAGCACTGTGATTCATTTGGAGCAAAAGAAAGAACTTGAAGGTAC 1255
|||||
Qy 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle 299
|||||
Db 1256 AACAAAAATTTAGAAAGAACTTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCAATATT 1315
|||||
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
|||||
Db 1316 TCTATAGTGTCTCTTCCCTGCTGATCTATGCATCTTATGCTCTGGCTTCTGGTATGG 1375
|||||
Qy 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
|||||
Db 1376 ACCACCTTGGTCTCTCAGGGGAATATTCATTGGACAAGTACTCTACTGTATTTCTTCT 1435
|||||
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
|||||
Db 1436 GTATTAATTGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAT 1495
|||||
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
|||||
Db 1496 GCAAGAGAGCAGCTTATGAATCTTCAAGATAATTCATTAATGAAGCAAGTATGACAGC 1555
|||||
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
|||||
Db 1556 TATTGGAAGAGTGGCCACAAACCAAGATAATATTAAAGGAAATTTGGAATTCAGAAATGTT 1615
|||||
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
|||||
Db 1616 CACTTCAAGTATACCATCTCGAAAAAGAGTAAAGATCTTGAAGGCGCTGAACCCGAGAGTG 1675
|||||
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
|||||
Db 1676 CAGAGTGGCAGACGGTGGCCCTGTGGTGGAAACAGTGGCTGTGGGAAGAGACACACAGTC 1735
|||||
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
|||||
Db 1736 CAGCTGATGACAGAGGCTCTATGACCCACAGAGGGGATGCTGCTGTTGTGACACAGGAT 1795
|||||
```

QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
|||||
Db 1796 ATTAGCACCATAAAGTAAAGTTCTACGGGAATCATTTGTTGGTGAGTCAGGAACCT 1855
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
|||||
Db 1856 GTATTGTTTGGCACCACGATAGCTGAAACATTCGCTATGGCCGTGAAATGTCCACCATG 1915
QY 500 AspGluIleGluLysAlaValIleGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
|||||
Db 1916 GATGAGATTGAGAAGCTGTCAAGGAAGCCCAATGCTATGACTTTATCATGAAACTGCCT 1975
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
:::|||||
Db 1976 CATAAATTTGACACCTGGTTGGAGAGAGGGGCCCAAGTTGAGTGGTGGGCAGACGAG 2035
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
|||||
Db 2036 AGGATCGCATTTGCACGAGCAAGCAAGCAGTGGTTCAAGTGGCTCTGGATAAGGCCGAGA 2155
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
|||||
Db 2156 AAAGTCGGACCCACCATTTGATAGCTCATCGTTGTCTACAGTTCGTAAATGCTGACGTC 2215
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
|||||
Db 2216 ATCGCTGGTTTCGATGATGGAGTCAATTTGGAGAAAGAAATCATGATGAATCATGAAA 2275
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
|||||
Db 2276 GAGAAAGCATTTACTTCAAACTTGTCAAAATGCAGACAGCAGGAAATGAAGTTGAATTA 2335
QY 640 GluAsnAlaThrGlyLysSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
|||||
Db 2336 GAAATGCGAGCTGATGATCCAAAGTGAAATTCATGCTTGGAAATGCTTCAATGAT 2395
QY 660 SerGlySerSerLeuIleLysArgSerThrArgSerIleHisAlaProGlnGly 679
||| |||||
Db 2396 TCAAGATCCAGCTCAATAAGAAAGATCAACTCGTAGGAGTCCCGTGGATCACAGACC 2455
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
|||||
Db 2456 CAAGCAGAAAGCTTAGTACCAGAGGCTCTGGATGAAGATACCTCCAGTTTCCCTTT 2515
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
|||||
Db 2516 TGGAGGATTATGAAGCTAAATTTAACTGAATGGCTTATTTTGTGTTGGTATTTGT 2575
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleGly 739
|||||
Db 2576 GCCATTATAATGAGGCTCCCAACAGCATTTTGCATAATAATATTTTCAAAAGATTATAGG 2635
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
:::|||||
Db 2636 GTTTTACAGAATTGATGATCCCTGAACAAACAGCAGAAATAGTAATCTGTTTCTACTA 2695
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
|||||
Db 2696 TTGTTTCTAGCCCTTGAATATTTCTTTATTTACATTTTTCCTTCAGGGTTTCACATTT 2755
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
|||||
Db 2756 GGCAAAAGTGGAGAGATCCTCACCAGGGCTCCGATACATGTTTCCGATCCATCCTC 2815
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
|||||
Db 2816 AGACAGGATGTGATGTTGTTGATGACCTTAAACACCACTGGAGCATTTCACTACAGG 2875
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
|||||

Db 2876 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTCCAGGCTGCTGTAATTACC 2935
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
|||||
Db 2936 CAGAAATATAGCAAAATCTTGGGACAGGAATAATATATATCTCTATCTATGTTGGCAACTA 2995
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
|||||
Db 2996 ACCTGTTACTCTTAGCAATTTACCCATCATTTGCAATAGCAGGATGTTGTAATGAA 3055
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
|||||
Db 3056 ATGTTGCTCGACAACACCTGAAAGATAGAAGAACTAGAAAGTCTGGGAAGATCGCT 3115
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
|||||
Db 3116 ACTGAAGCAATAGAAAACCTTCCGAACCGTTGTTTCTTCTTACTCAGGAGCAGAAAGTTGAA 3175
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
:::|||||
Db 3176 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATC 3235
QY 940 PheGlyValSerSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 3236 TTTGGAATTACATTTTCCCTTCCAGGCAATGATGATTTTCTTATGCTGGATGTTTC 3295
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
|||||
Db 3296 CGGTTGGAGCCYACTTGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 3355
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
|||||
Db 3356 TTTTACGCTGTTGCTTTGTCCTGCGCGTGGGCAAGTCAAGTTTCATTTGCTCTCTGAC 3415
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGlyLysSerProLeu 1019
|||||
Db 3416 TATGCCAAAGCCAAATATATCAGCAGCCACATCATCATTCATTAACAAAACCCCTTG 3475
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
|||||
Db 3476 ATTGACAGCTTACAGCAGGAAAGCCCTAATGCGCAACACATTTGGAAGAAATGTCATATT 3535
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
|||||
Db 3536 GGTGAAGTTGATTCAACTATCCACCACCGAGCATCCAGTGTCTCAGGAGACTGAGC 3595
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
|||||
Db 3596 CTGGAGTGAAGAAGGCCAGACGCTGGCTCTGGTGGCAGCAGTGGCTGTGGGAAGAC 3655
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
|||||
Db 3656 ACATGGTCCAGCTCTCGAGCGGTTCTACGACCCCTTGGCAGGAAAGTGTGCTGTGAT 3715
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
|||||
Db 3716 GGCAAGAAATAAAGCGACTGAATGTTAGTGGCTCCGAGCACACCTGGGCAATCGTGTCC 3775
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
|||||
Db 3776 CAGGAGCCCATCTGTTGACTGCACATTTGCTGAGAACATTCCTATATGAGAACACAGC 3835
QY 1140 ArgValValSerHisGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
|||||
Db 3836 CGGCTGCTGTACAGGAGAGATCGTGGGCGCAGCAAGAGGCGCAACATACATGCTCTTC 3895
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
|||||
Db 3896 ATCGATGCTACTGCTTAATAATATAGCTAAAGTAGGAGCAAAAGAACTCAGCTCTCT 3955
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
|||||

```
Db 3956 GGTGGCCAGAAAACAGCATTCGCCATAGCTGCTGGCCCTGTTAGACAGCCCTCATATTTTG 4015
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGGATGAAGCCACGCTCAGCTGGATACAGAAAGTGAAGAGTTGTCCAGAAGCC 4075
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTTGACAAAGCCAGAGAGCCGACCTGCATGTGATGTCTCACCGCTGTCACCATC 4135
Qy 1240 GlnAsnAlaAspLeuValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGAATGCAGACTTAATAGTGTGTTTTCAGAAATGCGAGAGTCAAGAGCAGTGGCAGCAT 4195
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleThrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTCTGGCAGAAAGGCATCATTTTTCATGGTCAAGTGTCCAGCTGGAACA 4255
Qy 1280 LysArg 1281
Db 4256 AAGCGC 4261

RESULT 15
AX336708
LOCUS AX336708 4646 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 7217 from Patent WO0194629.
ACCESSION AX336708
VERSION AX336708.1 GI:18127427
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrihan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 7217 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source 1..4646
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1371 a 892 c 1129 g 1254 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 4646
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 6 Gaps: 4

US-09-672-725c-25 (1-1281) x AX336708 (1-4646)
Qy 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheThrLysMetGly 19
Db 425 ATGGATCTTGAAGGGGACCGCAATGGAGCAGCAAGAAAGAACTTTTAAACTGAAC 484
Qy 20 LysLysSerLysLysGlyLysGlyLysGlyLysLysProThrValSerThrPheAlaMet 39
Db 485 AATAAAAGT---GAAAAGATAAGAAGGAAAGAAACCAACTGTCTCAGTGTATTTCAATG 541
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 542 TTTTCGTTATCAATTTGGCTTGACAAAGTTGTATATGGTGGGGAACCTTTGGCTGCCATC 601
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 602 ATCCATGGGGCTGGACTTCTCTCATGATGCTGGTGTGAGAGAAATGACAGATATCTTT 661
```

```
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 662 GCAAATGCAGGA---AATTTTGAAGATCTGATGTCAAACATCACTAATAGAGCTGATATC 718
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 719 AATGATACAGGGTCTTCATGAAT---CTGGAGGAGACATGACCAAGGATAGCTTATAT 775
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 776 TACAGTGAATGGTGTGGGGTCTGGTGTGCTGTACATTGAGTTTCATTTTGGTGC 835
Qy 140 LeuAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 836 CTGCGACGCTGGAAGACAAATACACAAATTAGAAAACAGTTTTTTTCATGCTATAATGCGA 895
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGlyLeuAsnThrArgLeuThrAsp 179
Db 896 CAGGAGATAGCTGGTGTGATGTGCAGATGTTGGGGAGCTTAACACCCGACTTACAGAT 955
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 956 GATGCTCTAAGATTATGAAGTTATTGGTGACAAAATTGGAATGTTCTTTCAAGTCAATG 1015
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 1016 GCAACATTTTTCACCTGGGTTTATAGTAGATTTACACGTGGTTGGAGCTTAACCCCTGTG 1075
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 1076 ATTTTGGCCATCAGCTCTGTTCTTGGACTGCTCAGCTGCTCTGGGCAAGATACATCTCT 1135
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 1136 TCATTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTC 1195
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 1196 TTGCGACAAATTAGAACTGTGATTCATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 1255
Qy 280 AsnLysAsnLeuGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle 299
Db 1256 AACAAAATTTAGAAAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCAATATT 1315
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 1316 TCTATAGTGTGCTTCTCTGCTGATCATGCTATGCTCTGCTGCTGCTGCTGCTGCTGCTG 1375
Qy 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 1376 ACCACCTGGTCTCTCAGGGGAATATTCATTGGGCAAGTACTACTGTTCTTTCTTCT 1435
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1436 GTATTATTTGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAT 1495
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1496 GCAAGAGGAGCAGCTTATGAAATCTTCAAGATAATTGATAATAGCCAACTATTGACAGC 1555
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1556 TATTTCGAAGAGTGGGCAACCAACAGATAATATTAAAGGAAATTTGGAAATTCAGAAATGT 1615
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1616 CACTTCAGTTACCCATCTCGAAAAGAAAGTTAAGATCTTTGAAGGGCCTGAACTGAAGGTG 1675
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1676 CAGAGTGGCAGACGGTGGCCCTGTTGGAAACAGTGGCTGTGGGAAAGACACAAACAGTC 1735
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
```

Db 1736 CAGCTGATGACAGAGCTCTATGACCCACAGAGGGGATGGTCAGTGTTCATGGACAGGAT 1795
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1796 ATTAGACCATTAATGTAGGTCTACGGGAATCATTTGGTGTGGTGGAGCAAGCCT 1855
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1856 GTATTGTTGCCACCACGATAGCTGAAACATTCGCTATGCGGTGAAATGTCCACCATG 1915
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1916 GATGAGATTGAGAAAGCTGTCAAGGAAGCCCAATGCGCTATGATCTTATCATGAAACTGCCT 1975
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1976 CATAAATTGACACCTGGTGGAGAGAGAGGGGCCAGTGTGAGTGGGCGAAGCAG 2035
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 2036 AGGATCGCATGACAGTGCCTGTTCCGAACCCCAAGATCCTCCTGCTGGATGAGGCC 2095
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArg 579
Db 2096 ACGTCAGCCTTGGACACAGAAAGCAAGCAGTGTTCAGGTGGCTCTGGATAAGGCCAGA 2155
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 2156 AAGTCGGACCACTTGTGATAGCTCATCTGTTGTCTACAGTTCGTAATGCTGACGTC 2215
QY 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 2216 ATCGCTGTTTCGATGATGGAGTCATTGTGGAGAAAGCAATCATGATGAATCATGAA 2275
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 2276 GAGAAAGCATTTACTTCAAACTTGTCAAAATGTCACAAATGCAGACAGCAGGAAATGAAGTTGAATTA 2335
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2336 GAAATGCGAGCTGATGAATCCAAAGTGAATTTGATGCTTGGAAATGCTTCAAAATGAT 2395
QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2396 TCAAGATCCAGTCTAATAAGAAAGATCAACTCGTAGAGTGTCCGTGGATCACAGCC 2455
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
Db 2456 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGATATACCTCCAGTTTCCTTT 2515
QY 700 TrpArgIleLeuLysLeuAsnSerThrGlnTrpProTyrPheValValGlyIlePheCys 719
Db 2516 TGGAGATTATGAGCTAAATTTAACTGAATGGCCTATTATTGTTGGTGTATTGTTGT 2575
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2576 GCCATTATAATGAGGCTCGCAACCAAGCAATTTGCCAATAATATTTTCAAGATTTATAGG 2635
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 GTTTTACAGAAATGTGATGATCTCTGAAACAAACGACAGATACTAATCTGTTTCTACTA 2695
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2696 TTGTTTCAGCCCTTGGAAATTTATTTTATTTATCATTTTTCCTTCAGGGTTTCACATTT 2755
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 GGCAAGCTGGAGAGATCTCTACCAAGGGCTCCGATACATGGTTTCCGATCCATCCTC 2815
QY 800 ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819

Db 2816 AGACAGGATGTGAGTGTGGTGTGATGACCCCTAAAAACACCACCTGGAGCATTTGACTACCAGG 2875
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2876 CTCGCCAATGATGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGTCTGTAATTACC 2935
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTyrGlnLeu 859
Db 2936 CAGAAATATAGCAATCTTGGACAGGAATAATATATATCTTCATCTATGCTTGGCAACTA 2995
QY 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2996 ACATGTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGGAGTGTGTGAAATGAA 3055
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 3056 ATGTTGCTCGCAGACAGCACTGAAGATAGAAGAACTAGAAGTGTCTGGGAAGATCGCT 3115
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 3116 ACTGAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAAAGTTGAA 3175
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGTATGCTCAGAGTTTTCAGGTACCATACAGAACTCTTTTGAAGAAAGCACACATC 3235
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTGGAATTCATTTTCTTCCACCCAGGCAATGATGATTTTTCCTATGCTGATGTTTC 3295
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 3296 CGGTTTGGAGCCTACTTGGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 3355
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3356 TTTTCAGCTGTGCTTGTGTCATGGCGTGGGCAAGTCAAGTTCATTTGCTCCTGAC 3415
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TAGCCAAAGCCAAATATCAGCAGCCCATCATCATGATCATTAATAAAGCCCTTTG 3475
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTGACAGCTACAGCAGGAGGCCTAATGCCGACACACATTTGGAAGAAATGTCACATTT 3535
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGATTCAACTATCCACCCGACCGAGCATCCAGTGTCTCAGGAGCTGAGC 3595
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3596 CTGGAGTGAAGAAAGGCCAGACGCTGGCTCTGGTGGGACAGCTGGCTGTGGGAAGAGC 3655
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3656 ACAGTGGTCCAGCTCTCGGAGCGGTCTACGACCCCTTGGCAGGGAAGTGTGCTGTAT 3715
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 GGCAAGAAATAAAGGAGCTGAATTTCACTGGCTCCGAGCACACCTGGGCATCGTGTCC 3775
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 CAGAGCCCATCTCTGTTGACTGCAGCATTTGCTGAGAACTTGCCTTATGGAGACACAGC 3835
QY 1140 ArgValValSerHisGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
Db 3836 CGGGTGGTGCACAGGAGAGATCTGTAGGGCAGCAAGAGGCCAACATACATGCCTTC 3895
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCGAGTCACTGCTCTAATAATATAGCTAAAGTAGGAGACAAAGAACTCAGCTCTCT 3955

```
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GGTGGCCAGAAACACGCATTGCCATAGCTCGTGGCCCTTTGTAGACAGCCTCATATTTTG 4015
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGTGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAAAGTTGTCCAAGAAGCC 4075
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGGACAAAGCCAGAGAAAGCCGACCTGCATTGTGATTGCTCACCGCCTGTCCACCATC 4135
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGAAATGCAGACTTATATAGTGTGTTTCAGATGGCAGAGTCAAGAGGATGGCAGGCAT 4195
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 4255
QY 1280 LysArg 1281
Db 4256 AAGCGC 4261
```

Search completed: November 7, 2002, 05:04:39
Job time : 4106.81 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:36:49 ; Search time 301.645 Seconds
(without alignments)
7291.243 Million cell updates/sec

Title: US-09-672-725C-25

Perfect score: 6473

Sequence: 1 MDPEGGRKGSAAKFNWKGK.....LLAQKGIYFMSVQAGAKR 1281

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=Cygn2_1/USPTO_spool/US09672725/runat_04112002_124338_17877/app_query.fasta_1.6435
-DB=N_Geneseq_032802 -QMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725.cgn_1.1.0 @runat_04112002_124338_17877 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	6473	100.0	4279	22	AAD03505 Dog P-glycoprotein
2	6468	99.9	4279	22	AAD03504 Dog P-glycoprotein
3	6466	99.9	4279	22	AAD03506 Dog P-glycoprotein
4	6463	99.8	4279	22	AAD03488 Dog P-glycoprotein
5	6425.5	99.3	4317	22	AAD03489 Dog P-glycoprotein
6	5866	90.6	3860	21	AZ49332 Human wild-type mu
7	5866	90.6	3860	24	ABA94365 Human BCRP DNA rel
8	5862	90.6	4186	22	AAF86127 Cynomolgous monke
9	5859.5	90.5	4195	22	AAF86128 Cynomolgous monke
10	5857	90.5	3860	21	AZ49333 Human G185V mutant
11	5857	90.5	3860	24	ABA94366 Human BCRP DNA rel
12	5857	90.5	4349	22	AAH57442 Human intestine ce
13	5857	90.5	4646	21	AZ94738 Human ATP binding
14	5857	90.5	6505	17	AAT13394 Hybrid vector pSF-
15	5857	90.5	8630	21	AZ24041 Retroviral M4 mdr-
16	5857	90.5	8630	21	AZ24042 Sequence of human
17	5854	90.4	4669	8	AAH70752 Multidrug Resistan
18	5851	90.4	4378	11	AAQ04522 Human MDR-1 DNA.
19	5849.5	90.4	3988	21	AZ88973 Human multidrug re
20	5844	90.3	4646	15	AAQ72872 Human multidrug re
21	5840	90.2	4669	14	AAQ52726 Sequence of human
22	5830	90.1	4669	19	AAV32645 Human P-glycoprote
23	5815.5	89.8	4264	19	AAV66533 Mutated human P-gl
24	5815.5	89.8	4264	19	AAV66534 Murine multidrug r
25	5641	87.1	4788	21	AZ49335 Mouse BCRP DNA rel
26	5641	87.1	4788	24	ABA94368 Rat multidrug resi
27	5621	86.8	4425	21	AZ52048 Rat multidrug resi
28	5611	86.7	4369	21	AZ52047 Murine multidrug r
29	5301	81.9	4189	21	AZ49334 Mouse BCRP DNA rel
30	5301	81.9	4189	24	ABA94367 Mouse multidrug re
31	5301	81.9	4313	14	AZ38950 Rat mdr1b2 (multis
32	5291.5	81.7	4233	21	AZ90198 Rat mdr1b2 (multis
33	5291.5	81.7	4233	22	AAF27498 Human ATP binding
34	4915.5	75.9	3924	21	AZ94742 Human MDR-3 DNA.
35	4915.5	75.9	3924	21	AZ88974 Multidrug-resistan
36	3628	56.0	2726	15	AAQ70916 Multidrug-resistan
37	3628	56.0	2726	15	AAQ70916 Multidrug-resistan
38	3628	56.0	2726	18	AAT43322 Human ATP binding
39	3302.5	51.0	4776	21	AZ94744 H. contortus PGP-A
40	2822.5	43.6	4175	20	AAV69392 Drosophila melanog
41	2694	41.6	3942	23	ABL14175 Drosophila melanog
42	2555	39.5	4157	23	ABL06835 cDNA encoding the
43	2513	38.8	4047	20	AAH60201 Drosophila melanog
44	2491	38.5	4111	23	ABL16205 cDNA encoding mult
45	2481	38.3	4002	20	AAH5823

ALIGNMENTS

RESULT 1	
AAD03505	
ID	AAD03505 standard; cDNA; 4279 BP.
XX	
AC	AAD03505;
XX	
DT	13-JUN-2001 (first entry)
XX	
DE	Dog P-glycoprotein (PGP) allelic variant (Genotype B) cDNA.
XX	
KW	Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
KW	MDRI; drug bioavailability; transgenic animal; genetic model; ss.
XX	
OS	Canis familiaris.
XX	
FH	Key Location/Qualifiers
CDS	17..3862
FT	/*tag= a

Qy	441	LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	460
Db	1337	CTGATGCAGAGGGTCTATGACCCACAGATGCATGGTCTGTATTGATGGACAGACATT	1396
Qy	461	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	480
Db	1397	AGGACCAATAATGTAGAGCATCTTCGGGAAATTACTGGTGTGGTGAGTCAGAGGCCGTGTG	1456
Qy	481	LeuPheAlaThrThrIleAlaGluAsnIleLeuArgTyrGlyArgGluAsnValThrMetAsp	500
Db	1457	TTGTTTCCACCACGATAGCTGAACACATTCGCTATGGCCGCGAAATGTCCACATGGAT	1516
Qy	501	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	520
Db	1517	GAGATTGAGAAAGCTGTTAAGGAAGCAATGCCTATGATTTTATCATGAAACTACCTAAT	1576
Qy	521	LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg	540
Db	1577	AAATTTGACACTCTGGTGTGAGAGAGAGGGGCCAGCTGAGTGGTGGACAGAAACAGAGA	1636
Qy	541	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	560
Db	1637	ATCCCACTTCCTGGGCCCTGGTTCGCAACCCACAGATTCCTCTGCTGGATGAGCAACG	1696
Qy	561	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	580
Db	1697	TCAGCTCTGACACTCAAAAGTGAAGCAGTGGTTCAGGTGGCCCTGGATAAGGCCAGAAAA	1756
Qy	581	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	600
Db	1757	GGCCGGACTACCAATGTGATAGCTATCGTTTGTCTACAGTTCGTAATGCCGATGTCAAT	1816
Qy	601	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	620
Db	1817	GCTGTTTTCATCATGGAGTCATCTGTGGAAAGAAATCATGATGAACATCATGAAGAAG	1876
Qy	621	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	640
Db	1877	AAGGGCATTTACPTCAAACTGTGCAATGTCAGACAAGAGAAATGAAATGAGTTAGTAAG	1936
Qy	641	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	660
Db	1937	ATGCCCACTGGTGAATCCAAAAGTGAAGTATGGTCTGGAAATGTCTCCAAAAGATTC	1996
Qy	661	GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln	680
Db	1997	GGGTCCAGTTTAATAAAGAGATCACTCGCAGGAGTATACATGCACCAACAGGCCAA	2056
Qy	681	AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp	700
Db	2057	GACAGAAAGCTGGTACAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCTCTCTGG	2116
Qy	701	ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla	720
Db	2117	AGGATCTGAGCTGAACCTCAACTGAATGGCCCTATTGTGGTGTGATATTTTGTGCT	2176
Qy	721	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	740
Db	2177	ATTATAACGGAGGCTCGAACCAACATTTCAATAATATTTTCCAGAGATTAAGGATC	2236
Qy	741	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	760
Db	2237	TTTACCGGAGATGAGGATCCTCAAAACAAACACAGAAATAGTAACATGTTTCTGTATTG	2296
Qy	761	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	780
Db	2297	TTTCTAGTCTCTGGAAATATTTCTTTTATTAATCATTTTCTCCAGGCGTTACATTTGGC	2356
Qy	781	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	800
Db	2357	AAAGCTGGGGAGNCTCTACTAAGCGGCTTCGATACATGGTTTTCAGATCCATGCTGAGA	2416
Qy	801	GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	820

Db	2417		CAGGATGTCAGCTGGTTTGATGACCCCTAAAAACACCACCTGGAGCATTGACAACCAGGCTT	2476	
Qy	821		AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	840	
Db	2477		GCCAAATGATGGGCTCAAGTTAAAGGGGCTATAGSTTCACAGCTTGCTGTCAATACCCAG	2536	
Qy	841		AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleYrGlyTrpGlnLeuThr	860	
Db	2537		AATATAGCAATCTTGGACAGGCATATTATATCTTAATCTATGTTGGCAATTAAACA	2596	
Qy	861		LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet	880	
Db	2597		CTTTTACTCTTAGCAATTGTACCCCATCTTGAATAGCAGGAGTTGTTGAAATGAAATG	2656	
Qy	881		LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr	900	
Db	2657		TTGTCTGGACAGCACTGAAAGATAGAAGAGCTAGAAAGAGCTGGGAAGATTGCTACA	2716	
Qy	901		GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluThr	920	
Db	2717		GAAGCCATCGAAACTTCCGAACTGTTGTTCTTTGACTCGGGACAGCAAGTTTGAATAC	2776	
Qy	921		MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	940	
Db	2777		ATGTATGCACAGAGTTTGGCAAGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTC	2836	
Qy	941		GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	960	
Db	2837		GGGTGCTCATTTTCTATCACCCAGGCAATGATGATATTTTCTATGCTGGCTGTTCCGG	2896	
Qy	961		PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe	980	
Db	2897		TTTGGTGCCTACTTGTGGCAATGAGTTCTATGAACTTTCAGGATGTTCTTTGGTATTC	2956	
Qy	981		SerAlaIleValPheGlyAlaMetalValGlyGlnValSerSerPheAlaProAspTyr	1000	
Db	2957		TCAGCTATTGCTTTGGTGCATGGCAGTGGGCGAGTCAGTTTCATTTGCTCCTGACTAT	3016	
Qy	1001		AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1020	
Db	3017		GCCAAAGCCAAAGATATCAGCAGCCACGTCATCATGATCATGTAAGAAAGCCCTCTGATT	3076	
Qy	1021		AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1040	
Db	3077		GACAGCTACAGCCCTCAGGCCCTCAGCCCAATATAGTTTGGAGGAAATGTGACATTTAAT	3136	
Qy	1041		GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1060	
Db	3137		GAGTCTGTTTCAACTATCCACTCAGCCAGACATCCCGCTGCTCCAGGGGCTCAGCCTC	3196	
Qy	1061		GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1080	
Db	3197		GAGGTGAAGAAGGGCCAGACGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGAGACA	3256	
Qy	1081		ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1100	
Db	3257		GTGTTCAGCTCCTAGAGGCTTCTATGACCCCTTGGCTGGTTCAGTGTCTAATGTATGGC	3316	
Qy	1101		LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1120	
Db	3317		AAAGAGATAAAGCACCTGAATGTCCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAG	3376	
Qy	1121		GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1140	
Db	3377		GAGCCATCTCTTTTACTCGAGCATTCGCGAAGACATTGGCTATGGAGACAACAGCCGG	3436	
Qy	1141		ValValSerHisGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle	1160	
Db	3437		GTGATACATGAAGAGATTATGACGGCAGCCAGGAGGCCACATACACCTTCATCTCATC	3496	
Qy	1161		GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1180	

Db 3497 GAGACACTCCCTGAGAAATACACACAGAGTAGGAGACAAAGAACCCAGCTCTCTGGT 3556

Qy 1181 GlyGlnLysGlnArgGlnAlaAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200

Db 3557 GGCAGAAACAGCGCATGCCATAGCTCCGCTCTGTGTAGACAGCCATATTTGGTT 3616

Qy 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220

Db 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAGAACCCCTG 3676

Qy 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240

Db 3677 GACAAAGCCAGAGAGCGCGACCTGATGTGATCGCCACCGCTGTGCCACCATCCAG 3736

Qy 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260

Db 3737 AATGCAGATTATAGTGTGTTCAGATGGCAAGTCAAGAGCATGGCAGACATCA 3796

Qy 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280

Db 3797 CAGCTGCTGCCCCAGAAAGGCATCTATTTTCATGTGTCCAGGTCCAGGCTGGAGCAAAG 3856

Qy 1281 Arg 1281

Db 3857 CGC 3859

RESULT 2

AAD03504

ID AAD03504 standard; cDNA; 4279 BP.

XX

AC AAD03504;

XX

XX

DT 13-JUN-2001 (first entry)

XX

DE Dog P-glycoprotein (PGP) allelic variant (Genotype A) cDNA.

XX

KW Dog; P-glycoprotein allelic variant; multidrug transporter; MDRL;

KW drug bioavailability; transgenic animal; genetic model; ss.

XX

OS Canis familiaris.

XX

FH Key

FD Location/Qualifiers

FT 17..3862

FT /*tag= a

FT /product= "dog P-glycoprotein (PGP) allelic variant

FT (Genotype A) protein"

FT replace (607, C)

FT /*tag= b

FT allele

XX

PN W0200123540-A2.

XX

XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-US26767.

XX

PR 28-SEP-1999; 99US-0156510.

XX

PA (GENT-) GENTEST CORP.

XX

PI Stocker PJ, Steinel-crespi DT, Crespi CL, Reif TC, Patten CJ;

XX

XX WPI; 2001-235373/24.

DR P-PSDB; AAE00308.

XX

XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful

PT for determining the bioavailability of drugs and for screening for dog

PT PGP inhibitors -

XX

XX Claim 9; Page 85-90; 111pp; English.

PS

XX The invention relates to dog P-glycoprotein (PGP) also referred

CC as multidrug transporter (MDRL) and nucleic acids encoding them.

CC The invention also includes fragments and biologically functional

CC variants of dog P-glycoprotein. PGP and their nucleic acids are

CC screening PGP inhibitors. They are useful for the diagnosis and

CC treatment of conditions characterised by PGP activity, by

CC reducing or increasing PGP activity in a cell. PGP nucleic acids

CC are used as oligonucleotide probes. Complements of PGP nucleic

CC acids are useful as antisense oligonucleotides, to induce a PGP

CC 'knockout' phenotype. They are used to prepare a non-human

CC transgenic animal, which are valuable as genetic models for

CC human diseases.

CC The present sequence is dog P-glycoprotein (PGP) allelic variant

CC (Genotype A) cDNA. The PGP enzyme functions as an efflux pump

CC exporting small molecules across the cell membrane. This enzyme

CC is a member of the ABC transporter family.

XX

SQ Sequence 4279 BP; 1295 A; 833 C; 1008 G; 1143 T; 0 other;

Alignment Scores:

Pred. NO.: 0 Length: 4279

Score: 6468.00 Matches: 1280

Percent Similarity: 99.92% Conservative: 0

Best Local Similarity: 99.92% Mismatches: 1

Query Match: 99.92% Indels: 0

DB: 22 Gaps: 0

US-09-672-725C-25 (1-1281) x AAD03504 (1-4279)

Qy 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20

Db 17 ATGGATCTCTGAAGAGCGCTAAGGGAGTGCAGAGAACTTCTGAAAATGGGCAA 76

Qy 21 LysSerLysLysLysGluLysLysLysProThrValSerThrPheAlaMetPhe 40

Db 77 AAAAGTAAAAAATGAGAGAAAGAAACCACTGTCCAGCAGTTTGCATGTTT 136

Qy 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 60

Db 137 CGCTATTCAAAATGGCTTGATAGTTGATATGTTGGTGGGACAATGGCTGCCATC 196

Qy 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80

Db 197 CATGGAGTGCACCTCCCTCTCATGATGCTGTTTTTGGAAACATGACAGATGCTTGA 256

Qy 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100

Db 257 AATGAGGAATTTCAAGAAACAACTTTCCAGTTATAATTAATGAAGATTATACGAAC 316

Qy 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120

Db 317 AATACACAACTTTCATCAACCATCTGGAGGAGGAATGACACGATGCTATTATTAC 376

Qy 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140

Db 377 AGTGGATCGGTGCTGGCTGCTGCTTACATCCAGGTTTCATTCTGGTGGCTG 436

Qy 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160

Db 437 GCAGCAGGAAGACAGATACTCAAAATTAGAAAACAATTTTTCATGCTATCATCGGAC 496

Qy 161 GluIleGlyTyrPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180

Db 497 GAGATTGCTGGTTGACGTCATGACGCTGGGAGCTTAACACCGGCTCACAGACGAT 556

Qy 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIleAla 200

Db 557 GTCTCCAAAATCAATGAAGGAATTTGGCGACAAAATTTGGAATGTTCTTCAATCAATAGCA 616

Qy 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220

Db 617 ACATTTTTCACCGGTTTATAGTGGGTTTACAGTGGTGGAAAGCTAACCTTGTGAT 676

Qy 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240

Db 677 TTGGCCATCAGCCCTGTTCTTGGACTTTCAGCCGCCACTCTGGCAAGATACTATCTCA 736
Qy 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu 260
Db 737 TTTACTGATAAGAACTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 796
Qy 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluAurGtyrAsn 280
Db 797 GCAGCAATCAGAACTGTGATTGCCCTTTGGAGGACAAAGAAAGAACTTGAAAGGTACAAC 856
Qy 281 LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
Db 857 AAAAATTTAGAAAGCTTAAAGAAATTTGGGATAAAGAAAGCTATCAGGGCCAACTTTCT 916
Qy 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrGlyThr 320
Db 917 ATTGGTGGCGTTCCTTATTTGATCTATGCATCATATGCTCTGGCTTTCTGGTATGGACC 976
Qy 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPheSerVal 340
Db 977 TCCTTGGTCCCTCCAGTGAATATCTATTGGACAAGTACTCACTGCTCTTTCTCTGTA 1036
Qy 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db 1037 TTAATTGGGGCTTTAGTATTGGACAGGCATCCCAAGCATTTGAAGCATTTGCAACGCCA 1096
Qy 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
Db 1097 AGAGGACAGCTTATGAATCTTCAAGATATTGCACATTAACCAACCATTTGACAGCTAT 1156
Qy 391 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1157 TCGAAGAGTGACATAAACACAGATAATTAAGGAAATTTGGAATTCAAAATGTTTCCAC 1216
Qy 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1217 TTCAGTTACCCCTTCTCGAAAGAAAGTTAAAGATCTTAAAGGGTCTCAACCTGAAAGTTTCA 1276
Qy 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1277 AGTGGCAGACAGTGGCGCTGTTGGACAGTGGCTGCCGGAGACAGACCGGTGCAG 1336
Qy 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db 1337 CTGATGCAGAGGCTCTATGACCCACACAGATGGCATGCTGTATTGATGGACAGGACATT 1396
Qy 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db 1397 AGGACCAATAATGTAAGGCATCTTCGGGAAATTAAGTGGTGGTGGTGGTGGTGGTGGT 1456
Qy 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1457 TTGTTGCCACCATAGCTGAAACATTCGCTATGGCCGCCGAAATATGTCACCATGGAT 1516
Qy 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1517 GAGATTGAGAAAGCTGTTAAGGAAGCAATGCCATGATGATTTATCATGAAACTACCTAAT 1576
Qy 521 LysPheAspThrLeuValGluAurGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
Db 1577 AAATTTGACACTCTGGTGTGGAGAGAGAGGGGCCAGCTGAGTGGTGGACAAACAGAGA 1636
Qy 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
Db 1637 ATGCCAATTCCTGGGCCCTTGGTTCGCAACCCCAAGATTCCTCTGCGATGAGGCAACG 1696
Qy 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1697 TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCTCGGTAAGGCCAGAAA 1756
Qy 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1757 GGCGGGACTACCATTTGTGATAGCTATCGTTTGTCTACAGTTTCTGTAATGCCGATGTCAAT 1816

Qy 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1817 GCTGGTTTTGATGATGAGTCATTGTGGAGAAAGGAAATCATGATGAATCATGAAAGAG 1876
Qy 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1877 AAGGCATTTTACTTCAAACTTGTCACAATGCACACAAGAGGAAATGAAATGAGTTAGAA 1936
Qy 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1937 AATGCCACTGGTGAATCCAAAAGTGAAGTGAATGCTTGGAAATGCTCCAAAAGATTCA 1996
Qy 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db 1997 GGTCTCCAGTTTAATAAAGAAGATCACTCGCAGAGATATACATGCACCAAGGCCAA 2056
Qy 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
Db 2057 GACAGAAAGCTTGGTACAAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCTTCTGG 2116
Qy 701 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db 2117 AGGATTCTGAAGCTGAACCTCACTGAATGGCCTTATTTTGTGGTTGGTATATTTTGTGCT 2176
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2177 ATTATAAACGGAGGCTGCACACGACATTTCAATATATTTTCAAGGATTTATAGGATC 2236
Qy 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2237 TTTTACCCGAGATGAGGATCCTGAAACAAAACGACAGAAATAGTAACATGTTTCTGTATTG 2296
Qy 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2297 TTTCTAGTCTCTGGAAATATTTCTTTATATACATTTTCTCCAGGGCTTCACATTTGGC 2356
Qy 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAAGCTGGGGAGATCCTCACTAAGCGCTTCGATACATCGTTTTCAGATCCATGCTGAGA 2416
Qy 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGGATGTGAGTGGTGTGATGACCCATAAAACACCACTGGAGCATTTGCAACACCAAGCTT 2476
Qy 821 AlaAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2477 GCCAATGATCGGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGTCTATTACCCAG 2536
Qy 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
Db 2537 AATATAGCAATCTTGGGACAGCATTTATATATCTTAACTATGTTGGCAATTAACA 2596
Qy 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTTACTCTAGCAATTCACCATCATGCAATAGCAGGAGTTGTTGAAATGAAATG 2656
Qy 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
Db 2657 TTGCTCGGCAAGCACTGAAAGATAGAAAGAGCTAGAAAGAGCTGGGAAGATTCTCTACA 2716
Qy 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2717 GAAGCCATCGAAAACCTTCGAAACTGTTGTTCTTTGCTCTGGGACAGAGTTTGAATAC 2776
Qy 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2777 ATGTATGCACAGATTTGCAAGTACCATACAGAACTCTTTGAGGAAAGACACATCTTC 2836
Qy 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2837 GGGGTCTCTATTTCTACCCAGGCAATGATGATATTTTCTATGCTGGTGTGTTCCGG 2896

Qy	961	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe	980
Db	2897	TTTTGGTGCCCTACTTGGTGCAAAATGAGTTCATGAACCTTCAGCATGTTCTTTTGGTATTC	2956
Qy	981	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	1000
Db	2957	TCAGCTATTGTTTGGTGCCATGCGAGTGGGCGAGGTCAGTTCATTGCTCCTGACTAT	3016
Qy	1001	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1020
Db	3017	GGCAAGCCAAAAGTATCAGACGCCACGCTCATCATCATTTGAAAAAGCCCTGATTT	3076
Qy	1021	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1040
Db	3077	GACAGCTACAGCCCTCACGCCCTCAAGCCAAATAGCTTGGAGGAATGTGACATTTAAT	3136
Qy	1041	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1060
Db	3137	GAGGTGCTGTTCAACTATCCCACTCGACCAGACATCCCGCTGCTCCAGGGGCTGAGCCTC	3196
Qy	1061	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1080
Db	3197	GAGGTGAAGAAGGCCAGACGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGACACA	3256
Qy	1081	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1100
Db	3257	GTGTTCAGCTCCTAGAGCGCTTCTATGACCCCTTGGCTGGTTCAGTGCTAATTCATGCG	3316
Qy	1101	LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1120
Db	3317	AAAGAGATAAGCACCTGAATGTCCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAG	3376
Qy	1121	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1140
Db	3377	GAGCCCATCTGTTTGACTGCGACATGTCCGAGAACTTGCCTRTGGAGACACACCGCG	3436
Qy	1141	ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle	1160
Db	3437	GTGCTATCATCATGAAGAGATTATGCAGGCAGGCCAAGGAGGCCCAACATACACCCTTCATC	3496
Qy	1161	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1180
Db	3497	GAGACACTCCCTGAGAAATACACACAGTAGGAGACAAAGTGAAGAGTGTCCCAAGAAGCCCTG	3556
Qy	1181	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1200
Db	3557	GGCCAGAAACAGCGCATTGCCATAGTCCGGCTCTGTGTAGACAGCCTCATATTTGCTT	3616
Qy	1201	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1220
Db	3617	TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTGTCCCAAGAAGCCCTG	3676
Qy	1221	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1240
Db	3677	GACAAAGCCAGAGAGGCCGACCTGCATGTGTATGCGCCACCGCTGTGCCACATCCAG	3736
Qy	1241	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1260
Db	3737	AATGCAGATTATAGTGGTGTTCAGATGGCAAGTCAAGAGCATGGCACACATCAA	3796
Qy	1261	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys	1280
Db	3797	CAGCTGCTGCCCGAAGAGGCATCTATTTTTCATGTGTCCAGTGTCCAGGCTGGAGCAAG	3856
Qy	1281	Arg	1281
Db	3857	CGC	3859
RESULT	3		
AA003506			
ID	AA003506	standard; cDNA; 4279 BP.	
XX			
AC	AA003506;		

XX	13-JUN-2001	(first entry)	
DT			
XX			
DE	Dog P-glycoprotein (PGP) allelic variant (Genotype D) cDNA.		
XX			
KW	Dog; P-glycoprotein allelic variant; PGP; multidrug transporter; MDR1; drug bioavailability; transgenic animal; genetic model; ss.		
XX			
OS	Canis familiaris.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	17..3862	
FT		/*tag= a	
FT		/product= "Dog P-glycoprotein (PGP) allelic variant (Genotype D) protein"	
FT	allele	replace (91, T)	
FT		/*tag= b	
FT	allele	replace (607, C)	
FT		/*tag= c	
FT	allele	replace (1001, T)	
FT		/*tag= c	
FT	allele	replace (3458, A)	
FT		/*tag= c	
XX	WO200123540-A2.		
PN			
XX			
PD	05-APR-2001.		
XX			
PF	28-SEP-2000; 2000WO-US26767.		
XX			
PR	28-SEP-1999; 99US-0156510.		
XX	(GENT-) GENTEST CORP.		
PA			
PI	Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;		
XX			
DR	WPI; 2001-235373/24.		
XX	P-PSDB; AAE00310.		
XX	New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful for determining the bioavailability of drugs and for screening for dog PGP inhibitors -		
PT			
PT			
XX	Claim 9; Page 102-107; 111pp; English.		
PS			
XX	The invention relates to dog P-glycoprotein (PGP) also referred as multidrug transporter (MDR1) and nucleic acids encoding them.		
CC	The invention also includes fragments and biologically functional variants of dog P-glycoprotein. PGP and their nucleic acids are useful for determining the bioavailability of drugs and for screening PGP inhibitors. They are useful for the diagnosis and treatment of conditions characterised by PGP activity, by reducing or increasing PGP activity in a cell. PGP nucleic acids are used as oligonucleotide probes. Complements of PGP nucleic acids are useful as antisense oligonucleotides, to induce a PGP 'knockout' phenotype. They are used to prepare a non-human transgenic animal, which are valuable as genetic models for human diseases.		
CC	The present sequence is dog P-glycoprotein (PGP) allelic variant (Genotype D) cDNA. The PGP enzyme functions as an efflux pump exporting small molecules across the cell membrane. This enzyme is a member of the ABC transporter family.		
XX			
SQ	Sequence 4279 BP; 1296 A; 833 C; 1009 G; 1141 T; 0 other;		
Alignment Scores:			
Pred. No.:	0	Length:	4279
Score:	6466.00	Matches:	1279
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	99.84%	Mismatches:	0
Query Match:	99.89%	Indels:	0
DB:	22	Gaps:	0

us-09-672-725c-25 (1-1281) x AAD03506 (1-4279)

QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
DB 17 ATGGATCTCTGAAGAGCGCGTAAGGGAGTCGACAGAGAACTTCTGGAAATGGGCAA 76
QY 21 LysSerLysLysGluLysGluLysLysProThrValSerThrPheAlaMetPhe 40
DB 77 AAAAGTAAAAAAGAGAGAAAGAAAGAAACCACTGTCACGACGTTGGCAATGTT 136
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 60
DB 137 CGCTATTCAAATTCGCTGATAGTGTGTATATGTTGGGGACAAATGGCTGCATCATC 196
QY 61 HisGlyAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
DB 197 CATGAGCTGCACCTCTCATGATGCTGTTTTTGGAAACATGACAGATAGCTTTGCA 256
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
DB 257 AATGCAGGAATTTCAAGAAACAAACTTTCCAGTTATATTAATGAAGTATTACGAAC 316
QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120
DB 317 AATACACAACATTTTCATCAACCATCTGGAGGAGAAATGACCAGTATGCTATTATAC 376
QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
DB 377 AGTGGATCGGTGCTGCGTGTCTTACATCCAGGTTTCATTTCTGGTGCGTG 436
QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
DB 437 GCACAGAGACAGATACTCAAAATTAGAAACAATTTTTCATGCTATCATCGACAG 496
QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
DB 497 GAGATTGGCTGGTTGACGTGCATGACGTTGGGAGCTTAACACCCGCTCACAGACGAT 556
QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
DB 557 GTCTCCAAATCAATGAAGGAATGGCGCAAAATTTGGAATGTTCTTTCAATCAATAGCA 616
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
DB 617 ACATTTTTCACCGGTTTATAGTGGGTTTACAGTGGTTGGAAGCTAACCCCTTGAT 676
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
DB 677 TTGGCCATCAGCCCTGTTCTTGGACTTTCAGCCGCCATCTGGCAAGATATCTTCA 736
QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu 260
DB 737 TTACTATGAAGAACCTTTGGGCTATGCAAAAGCTGGACAGTAGCTGAAGAAGCTTA 796
QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
DB 797 GCACCAATCAGACTGTGATGGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAAC 856
QY 281 LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
DB 857 AAAAATTTAGAAGAAAGTAAAGAAATTTGGGATAAAGAAAGCTATCACGGCCAACTTCT 916
QY 301 IleGlyAlaAlaPheLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
DB 917 ATTGGTCCGCTTCTTATTGATCTATGCATCATATATGCTGCTGCTTCTGGTATGGGACC 976
QY 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
DB 977 TCCTTGGTCTCTCCAGTGAATATACTATTGGACAGTACTACTGCTCTCTTCTCTGTA 1036
QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
DB 1037 TTAATTGGGGCTTTTAGTATTGGACAGGCAATCCCAAGCATTGAAGCAITTTGCAACGCA 1096

QY 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
DB 1097 AGAGGAGCAGCTTATGAAATCTTCAAGATAAATTTGACAATAAAACCAAGCATTCACAGCAT 1156
QY 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
DB 1157 TCGAAGAGTGGACATAAACACAGATATATTAGGGAAATTTGGAATTTCAAAATGTTTAC 1216
QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
DB 1217 TTCAGTTACCTTCTCGAAAGAAAGTTAAGATCTTAAGGGTCTCAACCTGAAGGTTTAC 1276
QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
DB 1277 AGTGGGCGACAGTGGCGCTGTTGGGAACAGTGGCTGCGGGAAGACGACCGCTGCA 1336
QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
DB 1337 CTGATGCAGAGGCTCTATGACCCACAGATGGCATGCTGTATTGATGCAGACGACATT 1396
QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
DB 1397 AGGACCATAAATGAAGGCATCTTCGGGAAATTTACTGTGTGTGAGTCAGGAGCTGTG 1456
QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
DB 1457 TTGTTGGCCACCACGATAGTGAACATTCGCTATGGCCGCAAAATGTCACCATGGAT 1516
QY 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
DB 1517 GAGATTGAGAAAGCTGTTAAGGAGCCATGCTATGATTTTATCATGAACACTACCTAAT 1576
QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
DB 1577 AAATTTGACACTCTGTTGGAGAGAGAGGGCCAGCTGAGTGGTGGACAGAAACAGAGA 1636
QY 541 IleAlaIleAlaArgAlaLeuValAlaArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
DB 1637 ATCGCATTTGCTCGGGCCCTGTTGCAACCCCAAGATTTCTCTGCTGGATGAGCAAG 1696
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
DB 1697 TCAGCTCTGGACACTGAAGACTGACAGTGGTTTCAGTGGCCCTGGATGAAGCCAGAAA 1756
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
DB 1757 GGCCGAGCTACCATTTGTGATAGCTCATCTGTTGCTACAGTTTCGTAATGCCGATGTCATT 1816
QY 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
DB 1817 GCTGGTTTTGATGATGAGTCAATTTGGGAGAAAGAAATCATGATGAATCATGAAGAG 1876
QY 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
DB 1877 AAGGCAATTTACTTCAAACTTGTCAATGCACAGACAGGAGAAATGAATTTGAGTTAGA 1936
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
DB 1937 AATGCCACTGGTGAATCCAAAAAGTAAAGTATGCTTGGAAATGCTCTCCAAAAGATTCA 1996
QY 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
DB 1997 GGGTCCAGTTTAAATAAAAGAAAGATCAACTCGCAGGAGTATACATGACCAAGGCCAA 2056
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
DB 2057 GACAGAAAGCTTGGTACAAAGAGGACTTGAATGAGATGATGCTCCAGTTTCTTCTGG 2116
QY 701 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
DB 2117 AGGATTTCAAGCTGAACCTCAACTGAATGGCCTTATTTTGGTGGTATATTTTGTGCT 2176


```
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyTle 740
DB 2177 ATTATAAACGGAGGCTGCAACACCATTTTCAATAATATTTCAGGATTATAGGATC 2236
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
DB 2237 TTTACCCGAGATGAGGATCCTGAAACAAACACAGAGATAGTAAACATGTTTCTGTATG 2296
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
DB 2297 TTTCTAGTCCTTGGAAATATTCTTTTATATATTTTCCCTCCAGGGCTTCACATTGGC 2356
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
DB 2357 AAAGCTGGGAGATCCTCACTAAGCGCTTCGATACATGGTTTTCAGATCCATGCTGAGA 2416
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
DB 2417 CAGGATGTCAGCTGGTTTGTATGACCTTAAACACCACTGGAGCATTGACAACAGGCTT 2476
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
DB 2477 GCCATGATCGCGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTGTCTATPACCCAG 2536
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
DB 2537 AATATAGCAAACTCTGGGACAGCATTTATATATCTTAATCTATGTTGGCAATTAACA 2596
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
DB 2597 CTTTACTCTTAGCAATTTGATGACCTATGCAATAGCAGGAGTTGTTGAATGAATAATG 2656
QY 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
DB 2657 TTGCTGGACAGCACTGAAAGATGAAGAAGAGCTAGAAGGAGCTGGGAGATTCTTACA 2716
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
DB 2717 GAAGCCATGAAAACTTCGAACTGTTGTTTCTTGTACTCGGAGCAAGAGTTTGAATAC 2776
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
DB 2777 ATGTATGACAGAGTTTGCAGTACCATACAGAAATCTTTTGGAGAAAGCACATCTTC 2836
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
DB 2837 GGGGTCTCTATTTCTATCACCCAGCAATGATGATATTTTCTCTATGCTGGCTGTTCCGG 2896
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
DB 2897 TTTGGTGCCTACTTGGTGCAATGAGTTTCATGAACTTTTCAGGATGTTCTTTTGTATTC 2956
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
DB 2957 TCAGCTATTGCTCTTGGTGCCATGGGAGTGGGCGAGTCCAGTTCATTGCTCTCTGACATP 3016
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
DB 3017 GCCAAAGCCAAAGTATCACAGCCCAAGTCATCATGATCATTTGAAAGAGCCCTCTGATT 3076
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
DB 3077 GACAGCTACAGCCCTCACGGCTCAAGCCAAATACGTTTGAAGGAAATGTGACATTTAAT 3136
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
DB 3137 GAGGTGCTGTAACTATCTCCACTCGACCAGACATCCCGTCTCAGGGGCTGAGCCCTC 3196
QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
DB 3197 GAGGTGAAGAGGCCAGCGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGAGCACA 3256
QY 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
```

```
DB 3257 GTTGTTCAGCTCTAGAGCGCTTCTATGACCCCTTGCTGCTGTTCAAGTCTAAATGATGC 3316
QY 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaAlaHisLeuGlyIleValSerGln 1120
DB 3317 AAAGAGATAAAGCACCTGAATGTCAGTGGCTCCGAGCACACCTGGCATCGTGTCTCAG 3376
QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
DB 3377 GAGCCCATCTCTGTTGACTGCAGCATTCGCGAGAACATTCCTTATGAGACAACAGCCGG 3436
QY 1141 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1160
DB 3437 GTCGTATACATGAAGAGATTGTCAGGCGCAAGAGAGCCCAACATACACCACTTCATC 3496
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
DB 3497 GAGACATCTCTGAGAATACACACCGAGTAGGAGACAAGAACCCAGCTCTCTGGT 3556
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
DB 3557 GGCCAGAAACAGCGCATTCGATAGCTCGCGCTCTTGTAGACAGCCTCATATTTGCTT 3616
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
DB 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGGTTGTCGAAGAGCCCTG 3676
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
DB 3677 GACAAAGCCAGAGAGCGCCGACCTGCATTTGTATGCCCCCGCTTGTCCACCATCCAG 3736
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
DB 3737 AATGAGATTAAATAGTGGTGTTCAGAAATGGCAAGTCAAGGAGCATGGCACACATCAA 3796
QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
DB 3797 CAGTGTGTCGCCCAAGAAAGCATCTATTTTCCATGTCAGTGTCCAGGCTGGAGCAAG 3856
QY 1281 Arg 1281
DB 3857 CGC 3859
RESULT 4
AAD03488
ID AAD03488 standard; cdna; 4279 BP.
XX AC AAD03488;
XX AC 13-JUN-2001 (first entry)
XX DE Dog P-glycoprotein (PGP) cdna #1.
XX KW Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
XX KW drug bioavailability; transgenic animal; genetic model; ss.
XX OS Canis familiaris.
XX FH Key Location/Qualifiers
XX CD 17..3862
XX FT /*tag= a
XX FT /product= "Dog P-glycoprotein (PGP) #1"
XX FN WO200123540-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26767.
XX PR 28-SEP-1999; 99US-0156510.
XX PA (GENT-) GENTEST CORP.
```


PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX
DR WPI: 2001-235373/24.
XX P-PSDB; AAE00303.
PT
PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
XX Claim 3; Page 58-63; 11pp; English.
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDR1) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) cDNA. This
CC sequence is also referred as Genotype C cDNA. The
CC PGP enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.
XX
SQ Sequence 4279 BP; 1294 A; 834 C; 1008 G; 1143 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6463.00 Matches: 1279
Percent Similarity: 99.84% Conservative: 0
Best Local Similarity: 99.84% Mismatches: 2
Query Match: 99.85% Indels: 0
DB: 22 Gaps: 0

us-09-672-725c-25 (1-1281) x AAD03488 (1-4279)

Qy 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 17 ATGATCTCTGAAGAGCCGCTAGGGAGTGCAGAGAACTCTCTGGAATGGGCAA 76
Qy 21 LysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
Db 77 AAAAGTAAAAAATGAGAGAAAGAAAGAAACCAACTGTCAGCAGCTTTGCAATCTT 136
Qy 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 60
Db 137 CGCTATTCAAATGGCTGATAGGTTGTATATGTTGGTGGGACAATGGCTGCATCATC 196
Qy 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
Db 197 CATGAGCTGCACCCCTCTCATGCTGCTGTTTTTGGAAACATGACAGATAGCTTGA 256
Qy 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 257 AATGAGAAATTCAGAAACAAACATTTTCCAGTTATATTAATGAAGATTAACGAAC 316
Qy 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120
Db 317 AATACACAACATTTTCATCAACCATCTGGAGGAGAAATGACCAGTATGCTATTATAC 376
Qy 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
Db 377 AGTGGATCGGTGCTGGCGTCTGCTTACATCCAGGTTTCATCTGGTGGCTG 436
Qy 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
Db 437 GCAGCAGGAGACAGACTCAAAATAGAAAAACAATTTTTCATGCTATCATCGACAG 496

Qy 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
Db 497 GAGATTGGCTGGTTGACGTGCATGCGTTGGGAGCTTAACACCGGCTCACACAGAT 556
Qy 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
Db 557 GTCTCCAAAATCAATGAAGGAATTGGCGACAAAATTTGGAATGTTCTTTTCACTCAATAGCA 616
Qy 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
Db 617 ACATTTTTCACCGGTTTATAGTGGGTTTACACGTTGGTGAAGCTAACCCCTTGTGATT 676
Qy 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
Db 677 TTGGCCCATCAGCCCTGTTCTTGAGCTTCAGCGCCCATCTGGGCAAGATATCTTCTCA 736
Qy 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
Db 737 TTTACTGATAAAGAACTCTTGCCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 796
Qy 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
Db 797 GCAGCAATCAGAACTGTGATTGCTTTGGAGGACAAAAGAAAGAACTTGAAAGGTACAAAC 856
Qy 281 LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
Db 857 AAAAATTTAGAGAAGCTAAAGAAATGGGATTAAGAAAGCTATCACGCCCAACATTTCT 916
Qy 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
Db 917 ATTGGTGGCGCTTCTTATTGATCTATGATCATATGCTTGGCTTCTGTTGTTATGGGACC 976
Qy 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340
Db 977 TCCTTGGTCTCTCCAGTAGTAATATCTATTGGACAAGTACTCACTGCTCTTTTCTGTGA 1036
Qy 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db 1037 TTAATTGGGCTTTTAGTATTGGACAGGATCCCAAGCATTTGAAGCATTTGCAACGCA 1096
Qy 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
Db 1097 AGAGGAGCAGCTTATCAAAATCTTCAAGATAAATTGACAATAAAACCAAGCATTCACAGCTAT 1156
Qy 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1157 TCGAAGAGTGGACATAAACCCAGATAATATAAGGGAATTTGGAATTCAAAATGTTTAC 1216
Qy 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1217 TTCAGTTACCTTCTCGAAAGAGTTAAGATCTTAAAGGCTCAACCTGAAGGTTTACAG 1276
Qy 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1277 AGTGGCAGACAGTGGCGCTGGTTGGGAACAGTGGCTGCGGGAAGAGCAGCAGCGTGCAG 1336
Qy 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db 1337 CTGATGCAGAGGCTCTATGCCCCACAGATGGCATGCTGTATTGATGCAGAGGACATTT 1396
Qy 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db 1397 AGGACATAAATTAAGGCATCTTCGGGAAATTTACTGTTGGTGGTGGTGGTGGTGGTGG 1456
Qy 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1457 TTGTTGGCCACCAGTAGCTGAAAACATTCGCTATGGCGCGGAAAATGTCACCATGGAT 1516
Qy 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1517 GAGATTGAGAAGCTGTTAAGGAAGCAATGCCTATGATGATTTTATCATGAACACTACCTAAT 1576

QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnArg 540
Db 1577 AAATTTTGACACTCTGGTTGGAGAGAGAGGGGCCAGCTAGTGGTGACAGAAACAGAGA 1636
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
Db 1637 ATGCCATTGCTCGGCCCTGGTTCCCAACCCCAAGATTCTTCTGCTGGATGAGGCAAG 1696
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1697 TCAGCTCTGACACTCAAAAGTGAAGCAGTGGTTCAAGTGGCCCTGGATAAGSCGAGAAA 1756
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1757 GGGCGACTACCATTTGTATAGCTCATCGTTTGTCTACAGTTCTGAATGCGCATGCTATT 1816
QY 601 AlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1817 GCTGGTTTTGATGATGGAGTCAATGTGGAGAAAGGAAATCATGATGAATCATGAAGAG 1876
QY 621 LysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1877 AAGGGCATTTACTTCAAACTTGTCAAATGCAGACAAGAGGAAATGAAATTCAGTTAGAA 1936
QY 641 AsnAlaThrGlyClnSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1937 AATGCCACTGGTGAATCCAAAAGTGAAGTGTATGCTTGGAAATGCTCCAAAAGATTCA 1996
QY 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db 1997 GGGTCCAGTTTTAATAAAGAGATCACTCCGAGAGTATACATGCACCAAGGCCAA 2056
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
Db 2057 GACAGAAAGCTTGGTACAAAAGAGACTTGAATGAGAATGTACCTCCAGTTTCCCTTCGG 2116
QY 701 ArgIleLeuLysLeuAsnSerThrGluTrpProThrPheValValGlyIlePheCysAla 720
Db 2117 AGAATTTCTGAAGCTGAACCTCAACTGAATGGCCTTATTTGTGGTATATTTTGTGCT 2176
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2177 ATTATAACGGAGGCTGCAACACGATTTTCAATAATATTTTCAAGGATATATAGGATC 2236
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2237 TTATCCCGAGATGAGATCCTGAAACAAAACGACAGAAATAGTAACATGTTTCTGTATTG 2296
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlnPheThrPheGly 780
Db 2297 TTCTTAGTCTCTGGAAATATTTCTTTTATTTACATTTTCTCCAGGGCTTCACATTTGCG 2356
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAAGCTGGGAGATCCTCNCTAAGCGCTTCGATACATGGTTTTTCAGATCCATGCTGAGA 2416
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGGATGTCAAGTGTGTTGATGACCTTAAACACCACTCGAGGCAATTGACAACCAAGGCTT 2476
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2477 GCCAATGATGGCGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCTGTATACCCAG 2536
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
Db 2537 ATATAGCAAAATCTTGGACAGCATTTATATATCTTAATCTATGTTGGCAATTAACA 2596
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTCTACTTTAGCAATTTGATCCCATCATTTGCAATAGCAGGAGTGTGTGAATGAAATG 2656
QY 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 900

Db 2657 TTGCTCGACAAGCACCTCAAAAGATAAAGAGAGCTAGAAGAGCTGGGAAGATTGCTACA 2716
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2717 GAAGCATCGAAAACATCCCGAACHTGTGTTCTTTGACTCGGAGACAGAGTTTGAATAC 2776
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2777 ATGTATGCACAGAGTTTGCAGAGTACCATACAGAAACTCTTTTGAGAAAGCACACATCTTC 2836
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2837 GGGTCTCATTTTCTATCACCCAGGCAATGATGATTTTCTATGCTGCTGTTCCGG 2896
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 980
Db 2897 TTTGGTGCCTACTTGTGGCAATGAGTTCATGAACTTTCAGGATGTTCTTTGGTATTC 2956
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlnValSerSerPheAlaProAspTyr 1000
Db 2957 TCAGCTATTGCTTTGTTGGTCCATGGCAGTGGGCGAGTTCAGTTTCATTTGCTCCTGACTAT 3016
QY 1001 AlalysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
Db 3017 GCCAAAGCCAAAGTATCAGCAGCCACGTCATCATGATCATTTGAAAAAGCCCTCTGATT 3076
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3077 GACAGCTACAGCCCTCAGGGCTCAAGCCAAATACGTTGGAAGAAATGTGACATTAAAT 3136
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3137 GAGTCTGTTTCAACTATCCACTCGACACACATCCCCGTGCTCCAGGGGCTGAGCCTC 3196
QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
Db 3197 GAGTGAAGAGGGCCAGACGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGAGACACA 3256
QY 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
Db 3257 GTTCTTCAGCTCTAGAGCGCTTCTATGACCCCTTGCTGGTTCAGTCTAATGTATGGC 3316
QY 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120
Db 3317 AAAGAGATAAAGCACCTGAATTCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAG 3376
QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
Db 3377 GAGCCCATCTGTTGACTGCAGCATTTGCCGAGAACATTGCCCTATGGAGACAACAGCCGG 3436
QY 1141 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1160
Db 3437 GTCGTATCATGATGAAGATATTATGCAGGACGCCAAGAGGCCAATACACCACTTCATC 3496
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
Db 3497 GAGACACTCCCTGAGAAATACAAACACCAGAGTAGGAGACAAGAAACCCAGCTCTCTGGT 3556
QY 1181 GlyClnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
Db 3557 GGCCAGAAACAGGCGCATGTCATAGCTGGCGCTCTTGTATAGACAGCCTCATATTTTGC 3616
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
Db 3617 TTGGATGAAGCTACACAGCTCTGGATACAGAAAGTGAAAGGTTGTCCAAGAGGCCCTG 3676
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
Db 3677 GACAAAGCCAGAGAGGCCGACCTGCTATGTCATGCCACCGCTTGTCCACCATCCAG 3736
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGlyThrHisGln 1260

Db 3737 AATGCAGATTTAATAGTGGTGTTCAGAAATGCAAGGAGCATGGCCACACATCAA 3796
 QY 1261 GlnLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
 Db 3797 CAGCTGCTGGCCCAAGAGGCATCTATTTTCCATGGTCCAGTGTCCAGGCTGGAGCAAAG 3856
 QY 1281 Arg 1281
 |||

Db 3857 CGC 3859

RESULT 5

ID AAD03489 standard; cDNA; 4317 BP.

XX AAD03489;

XX 13-JUN-2001 (first entry)

XX Dog P-glycoprotein (PGP) cDNA #2.

XX Dog; P-glycoprotein; PGP; multidrug transporter; MDRL;
 KW drug bioavailability; transgenic animal; genetic model; ss.

XX Canis familiaris.

XX Key Location/Qualifiers

XX CDS 70..3912

XX /tag= a

XX /product= "Dog P-glycoprotein (PGP) #2"

XX WO200123540-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26767.

XX 28-SEP-1999; 99US-0156510.

XX (GENT-) GENTEST CORP.

PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 DR WPI; 2001-235373/24.
 DR P-PSDB; AAE00304.

XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors -

XX Claim 1; Page 66-72; 11lpp; English.

XX The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDRL) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.

CC The present sequence is dog P-glycoprotein (PGP) cDNA. The
 CC PGP enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.

XX Sequence 4317 BP; 1293 A; 844 C; 1019 G; 1161 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 4317

Score: 6425.50 Matches: 1273
 Percent Similarity: 99.69% Conservative: 4
 Best Local Similarity: 99.38% Mismatches: 3
 Query Match: 99.27% Indels: 1
 DB: 22 Gaps: 1

US-09-672-725C-25 (1-1281) x AAD03489 (1-4317)

QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
 Db 70 ATGGATCTGAAGGAGCCCTAAAGGGAGTGCAGAGAGAACTCTGGAAAATGGGCAA 129
 QY 21 LysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
 Db 130 AAAAGT---AAAAAGAGAGAAAGAAAACCAACTGTCAGCAGCTTGGCAATGTTT 186
 QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 60
 Db 187 CGCTATTCAAATGGCTTGATAGTTGTATGTTGGTGGGACAATGGCTGCCATCATC 246
 QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
 Db 247 CATGGAGCTGCACCTCCCTCTCATGATGCTGTTTTGGAAAACATGACAGATAGCTTTGCA 306
 QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
 Db 307 ATGCAGGAATTTCAAGAAACAAACTTTCCAGATTATATAATTAAGAGTATTACGAAC 366
 QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 120
 Db 367 AATACACACATTTTCATCAACCATCTGGAGGAGAAATGACCACGTATGCCTATTATTAC 426
 QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
 Db 427 AGTGGGATCGGTGCTGGCGTGTGGTGGCTGTATCATCCAGGTTTCATCTGGTGCCTG 486
 QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
 Db 487 GCAGCAGGAAGACAGACATCTCAAAATTAGAAAACAATTTTTCATGCTATCATGCCACAG 546
 QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
 Db 547 GAGATTGGCTGGTTGACGTGCATGACGTTGGGAGCTTAACACCCGGCTCACAGACGAT 606
 QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
 Db 607 GTCTCCAAAATCAATGAAGGAATGGCGACAAAGTTGGAATGTCTTTCAATCAATAGCA 666
 QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
 Db 667 ACATTTTCCACCGGTTTATAGTGGGTTTACACCTGGTTGGAAGCTAACCCCTTGTGATT 726
 QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
 Db 727 TTGGCCATCAGCCCTGCTTGTGACCTTCAGCGCCATCTGGGCAAGATACTATCTTCA 786
 QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
 Db 787 TTTACTGATAAAGAACTCTTGGCCTATGCAAAAGCTGGAGCAGTACTGAAGAAGTCTTA 846
 QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
 Db 847 GCAGCAATCAGAACTGTGATTGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAAC 906
 QY 281 LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
 Db 907 AAAAATTTAGAGAAGCTAAAGAAATTTGGGATAAAGAAAGCTATCACGCCCAACATTCT 966
 QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
 Db 967 ATTGGTGGCGCTTTCTTATTGATCTATGATCATATGCTGCTGGCTTCTGGTATGGGACC 1026
 QY 321 SerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerVal 340

Db	1027	TGGTCTCTCCAGTGAATATATACTATTGGACAGGTACTCACTGCTCTCTTTCTGTGA	1086
Qy	341	LeuIleclAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	360
Db	1087	TTAATTGGGGCTTTAGTATTGGACAGGCATCCCAAGCATTTGAAGCATTTGCAAGCGCA	1146
Qy	361	ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	380
Db	1147	AGAGGAGCAGCTTATGAAATCTTCAAGTAAATTGACAATAAACCAAGCATTGACAGCTAT	1206
Qy	381	SerLysSerGlyHisLysProAspAsnLleLysGlyAsnLeuGluPheLysAsnValHis	400
Db	1207	TCGAAGAGTGGACATAAACACAGATAATATTAGGGAAATTTGGAATTCAAAAATGTTCCAC	1266
Qy	401	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	420
Db	1267	TTCAGTTTACCCTTCTCGAAAAGAGTTAAGATCTTTAAAGGCTCTCAACCTGAAGGTTCCAG	1326
Qy	421	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln	440
Db	1327	AGTGGCGACAGCTGGCGCTGGTTGGGAACAGTGGCTCGGGAGAGACGACGCGTCGAC	1386
Qy	441	LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	460
Db	1387	CTGATGCAGAGGCTCTATGACCCACAGATGCATGGTCTGTATTGATGGACAGGACATT	1446
Qy	461	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	480
Db	1447	AGGACCATAAATTAAGGCATCTCGGGAATTAAGTGTGTGGTGTGAGTCAGGAGCCGTGTG	1506
Qy	481	LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	500
Db	1507	TTGTTTGGCCACCAAGATAGCTGAAACATTCGCTATGGCCGCGAANAATGTCACCATGGAT	1566
Qy	501	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	520
Db	1567	GAGATTGAGAAAGCTGTTAAGGAAGCCAATGCCTATGATTTTATCATGAAACTACCTAAT	1626
Qy	521	LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg	540
Db	1627	AAATTTTGACACTCTGTTGCGAGAGAGGGCGCGCTGAGTGTGGCAGAGAAACAGAGA	1686
Qy	541	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	560
Db	1687	ATGCCCATTCCTCGGGCCCTGGTTCGCAACCCACAGATTCCTTCGTGGATGAGGCAACG	1746
Qy	561	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	580
Db	1747	TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGTGGCCCTGGATTAAGGCCAGAAAA	1806
Qy	581	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	600
Db	1807	GGCGGGACTACCAATTGTGATAGCTATCGTGTGCTACAGTTCGTAAATGGCGATGTGCTATT	1866
Qy	601	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	620
Db	1867	CTGTGTTTGATCATGGATGCTATTGTGGAGAAGGAANAATCATGATGAATCATGAAAGAG	1926
Qy	621	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	640
Db	1927	AAGGGCAATTTTACTTCAAACTTGTCACAATGCAGACAAGGAAATGAAATTTGAGTTAGAA	1986
Qy	641	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	660
Db	1987	AATGCCACTGGTGAAATCCAAAAAGTGAAGTGAATGCTTTGGAATTTCTCCAAAAAGATTCA	2046
Qy	661	GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln	680
Db	2047	GGGTCCAGTTTAAATAAAGAAGATCACTCCGACGAGATATACATGCACCAAGGCCAA	2106
Qy	681	AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTyr	700

```
QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
|||||
Db 3247 GAGGTGAAGAGGCCAGACGCTGCCCTCGTAGGTAGTGGCTGTGGGAAGACACA 3306

QY 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
|||||
Db 3307 GTTGTTCAGCTCCTAGAGCGCTTCATGACCCCTTGGCTGGTTCAGTCTAATGTATGGC 3366

QY 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120
|||||
Db 3367 AAAGAGATAAAGCACCTGAATGTCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAG 3426

QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAsnSerArg 1140
|||||
Db 3427 GAGCCCATCTGTTTACTGTCAGCATTCGCCAGACATTCGCTATGGAGACACAGCCGG 3486

QY 1141 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1160
|||||
Db 3487 CTCGTATCACATGAAGAGATTATGAGCGCAGCCAAAGGAGGCCAATACACCACTTCATC 3546

QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
|||||
Db 3547 GAGACATCCCTCGAATAATACACACCAAGTAGGAGACAAAGAACCCAGCTCTCTGGT 3606

QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
|||||
Db 3607 GCCCAGAAACAGCGCATTCGCATAGCTCGCGCTCTTGTAGACAGCCATATTTTGGCT 3666

QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
|||||
Db 3667 TTGGATGAAGCTACATCAGCTCGATGATACAGAAAGTGAAGAGTGTCCAAAGAGCCCTG 3726

QY 1221 AspLysAlaArgGluArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
|||||
Db 3727 GACAAAGCCAGAGAGCGCCCTCGATTGTGATCGCCCGCCCTGTCCACCATCCAG 3786

QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
|||||
Db 3787 AATGCAGATTAAATAGTGTCTTCAGAAATGGCAAGTCAAGAGCATGGGCACACATCAA 3846

QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
|||||
Db 3847 CAGCTGTGCTGCTGAGAAAGCATCTATTTTCCATGATCAGTGTCCAGGCTGGAGCAAG 3906

QY 1281 Arg 1281
|||
Db 3907 CGC 3909

RESULT 6
AAZ49332 standard; cDNA; 3860 BP.
XX
AC AAZ49332;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human wild-type multidrug resistance-1 (MDR-1) cDNA.
XX
KW Multidrug resistance; MDR-1; P-glycoprotein;
KW transmembrane efflux pump; haematopoietic stem cell; transduction;
KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW gene therapy; gene replacement; genetic defect; thalassaemia;
KW Gauchier's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW cytokine; wild-type; ds.
XX
OS Homo sapiens.
XX
FH Key
FH CDS 1..3843
FT /*tag= a
FT /product= "Human wild-type MDR-1 protein"
FT replace (553..555, GTT)
```

```
FT /*tag= b
FT /note= "cDNA sequence of G185V human mutant MDR-1 given
XX in AAZ49333"
PN WO9961589-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11825.
XX
PR 28-MAY-1998; 98US-0086988.
XX
PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Sorrentino B, Bunting K;
XX
DR WPI; 2000-072615/06.
XX
DR P-PSDB; AAY58186.
XX
PT Ex vivo expansion of hematopoietic stem cells transduced with a
PT sequence encoding human multidrug resistance-1, used for bone marrow
PT transplantation -
XX
XX Claim 10; Page 68-70; 113pp; English.
XX
CC This sequence represents cDNA encoding human wild-type
CC multidrug resistance protein MDR-1. MDR-1 is a transmembrane
CC efflux pump, responsible for the export of drugs from cells,
CC particularly cancer cells. Wild-type MDR-1 shows increased
CC resistance to etoposide and decreased resistance to vinca
CC alkaloids compared with a mutant form (AAV58187) where the Gly at
CC position 185 is replaced by Val. The invention relates to transducing
CC haematopoietic stem cells with nucleic acid encoding an MDR protein
CC and culturing the modified cells. The modified haematopoietic stem
CC cells are useful in bone marrow transplantation (to reconstitute
CC haematopoietic systems in patients who have undergone chemotherapy or
CC radiation therapy) and in ex vivo gene therapy of genetic defects in
CC cells derived from haematopoietic stem cells, e.g., thalassaemia,
CC Gauchier's disease, sickle cell anaemia or leukaemia. The modified
CC cells can also be used to identify factors involved in regulating
CC proliferation and differentiation in haematopoietic stem cells.
CC Haematopoietic stem cells that express MDR-1 will be protected against
CC chemotherapeutic agents, so can be engrafted while the patient is
CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC provides sufficient cells to permit standard biochemical analysis.
CC Overexpression of MDR-1 allows cytokine-driven expansion of
CC haematopoietic stem cells by at least 10-fold compared with a maximum
XX of 4-fold in known procedures.
XX
SQ Sequence 3860 BP; 1135 A; 746 C; 957 G; 1022 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 0 Length: 3860
Score: 5866.00 Matches: 1164
Percent Similarity: 95.09% Conservative: 55
Best Local Similarity: 90.80% Mismatches: 59
Query Match: 90.62% Indels: 4
DB: 21 Gaps: 4
```

US-09-672-725C-25 (1-1281) x AAZ49332 (1-3860)

```
QY 1 MetAspProGluGlyGlyArgLysGlySerAla--GluLysAsnPheTrpLysMetGly 19
|||||
Db 1 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAGAAAGAACTTTTAACTGAAC 60

QY 20 LysLysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
|||||
Db 61 AATAAAGT--CAAAAGATAGAGAGGAAGAACCAACTGTGAGTATTTTCAATG 117

QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
|||||
Db 118 TTTCGCTATTCAAAATTGGCTTGACAAAGTGTATATGGTGGTGGAACTTTGGCTGCATC 177
```

QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
|||||
Db 178 ATCCATGGGCTGGACTCTCTCATGATGCTGGTGTGGGAAATGACAGATATCTTT 237
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
|||||
Db 238 GCAAAATGCAGGA--AATTTAGAGATCTGATGTCACAAACATCACTAATAGAAAGTGATATC 294
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
|||||
Db 295 AATGATACAGGGTCTCTCATGAAT--CTGGAGGAAGACATCAACCATATCCCTATTAT 351
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
|||||
Db 352 TACAGTGGAAATTTGGTCTGGGGTGGCTGGTTCGCTTACATTCAGGTTTCATTTTGGTGC 411
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
|||||
Db 412 CTGGCAGCTGGAAGACAAATACACAAATTAGAAAAACAGTTTTTTCATGCTATAATGCGA 471
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
|||||
Db 472 CAGGAGATAGGCTGGTTGTATGTGCAGATGTTGGGGAGCTTAACACCCGACTTACAGAT 531
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
|||||
Db 532 GATGCTCTAAGATTAAAGAGTATTTGGTGACAAATTTGGAATGTTCTTTCAGTCAATG 591
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrVal 219
|||||
Db 592 GCAACATTTTTCACCTGGGTTATAGTAGGATTTACACGTGGTTGGAAGCTTAACCCCTGTG 651
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleThrAlaLysIleLeuSer 239
|||||
Db 652 ATTTGGCCATCATCTCTGTTCTTGGACTCTCAGCTGCTGGGCAAGATACATATCT 711
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
|||||
Db 712 TCATTTACTGATAAAGAACTTTAGCGTATGCAAAAGCTGGACAGTAGCTGAAGAGGTC 771
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
|||||
Db 772 TTGGCACAATATGAACCTGTGATTTGGAGGACAAAGAAAGAACTTTGAAGGTAC 831
QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
|||||
Db 832 AACAAAAATTTAGNAGAGCTAAAGAAATTTGGATAAAGAAAGCTATACAGCCATAT 891
QY 300 SerIleGlyAlaAlaPheLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
|||||
Db 892 TCTATAGTGTGCTTTCTCTGCTGATCTATGCATCTTATGCTCTGGCCCTTCTGGTATGG 951
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
|||||
Db 952 ACCACCTTGGTCCCTCAGGGGAATATCTATTGGACAAGTACTCACTGATTTCTTTCT 1011
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
|||||
Db 1012 GTATTANTGGGGCTTTTAGTGTGGACAGGCATCTCCAAGCATTTGAAGCATTTGCAAAAT 1071
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
|||||
Db 1072 GCAAGAGGAGCAGCTTATGAATCTTCAAGATAATTGATAATAAGCAAGTATTTGACAGC 1131
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
|||||
Db 1132 TATTTCGAAGAGTGGGCAACCAACCAAGATATTAAAGGAAATTTGGAAATTCAGAAATGTT 1191
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
|||||
Db 1192 CACTTCAGTTACCCATCTCGAAAAAGAGTTAAGATCTTTGAAGGCCCTGAACCTGAAGGTG 1251
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrVal 439

Db 1252 CAGAGTGGGCAGACGGTGGCCCTGTTGGAAACAGTGGCTGTGGAAAGACACAAACAGTC 1311
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
|||||
Db 1312 CAGCTGATGTCAGAGGCTCTATGACCCACAGAGGGGATGGTCACTGTGTATGGACAGAT 1371
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
|||||
Db 1372 ATTAGGACCATTAATGTAAGGTTTCTACGGGAATCATTTGTTGGTCAAGAACCT 1431
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
|||||
Db 1432 GTATTGTTTGCACCCACCATAGCTGAAACATTCGCTATGGCCGTGAAATGTCCACATG 1491
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
|||||
Db 1492 GATGAGATTGAGAAGCTGTCAAGGAAGCCATGCCATGACTTTTTCATGAAACTGCCT 1551
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
|||||
Db 1552 CATAAATTTGACACCTGTTGGAGAGAGAGGGCCAGATTGAGTGGTGGCAGAACGAC 1611
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
|||||
Db 1612 AGGATCGCATTTGCACGTGCCCTGTTGCAACCCCAAGATCCTCCTGCTGGATGAGGCC 1671
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
|||||
Db 1672 ACGTCAGCCTTGGACACAGAAAGCAAGCACTGGTTCAGTGGCTCTGGATAAGCCAGA 1731
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
|||||
Db 1732 AAAGTCGAGCACCACTGATAGCTCATCGTTGTTGTACAGTTGCTAATGCTGACGTC 1791
QY 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
|||||
Db 1792 ATCGCTGGTTCGATGATGGATCATTTGGAGAAAGAAATCATGATGAACCTCATGAAA 1851
QY 620 GlnLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
|||||
Db 1852 GAGAAGGCAITTTACTTCAAACTGTCACATGCACAGACAGCAGGAAATGAAGTTGAATTA 1911
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
|||||
Db 1912 GAAAAATGCAGCTGATGAATCCAAAAAGTGAATTTGATGCTTGGAAATGCTTCAAAATGAT 1971
QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
|||||
Db 1972 TCAAGATCCAGTCTAATAGAAAAAGATCACTCTAGAGAGTGCCTGGATCACAGCC 2031
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
|||||
Db 2032 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGTATATACCTCCAGTTTCCCTTT 2091
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
|||||
Db 2092 TGGAGGATTAAGAAGCTAAATTAATCTGAATGGCCCTTATTTGTTGGTGTATTTGT 2151
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
|||||
Db 2152 GCCATTATAAATGGAGGCTGCACACGCAITTTGCAATAATATTTTCAAGATATATAGG 2211
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
|||||
Db 2212 GTTTTTACAGAAATTTGATGATCCTCGAAACAAACACAGAAATAGTAACTTGTTTTCACTA 2271
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
|||||
Db 2272 TTGTTCTACGCTTGGAAATTTATTTTATACATTTTTCCTTCAAGGTTTCAATTT 2331
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
|||||

Db 2332 GCGAAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATAGGTGTTTCCGATCCATGCTC 2391
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2392 AGACAGGATGTCAGTTGGTTGATGACCCCTAAACACACCATGGAGCATTCACACAGG 2451
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2452 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGCTGTAATACC 2511
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleIleGlyTrpGlnLeu 859
Db 2512 CAGAAATAGCAAAATCTTGGACAGGAATAATATATCTCTCATCTATGTTGGCACTA 2571
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2572 ACACGTGTACTCTTAGCAATGTACCCCATCATTCGAATAGCAGGATGTTGAAATGAA 2631
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2632 ATGTTGCTGGACAAGCACTGAAAGATAAGAAAGAACTAGAAAGTGTCTGGGAAGATCGCT 2691
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2692 ACTGAAGCAATAGAAAACCTCCGAACCCGTTCTTTTGTACTCAGGAGCAGAGATTGAA 2751
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2752 CATATGATGTCAGAGTTTCAGGTACCATCACAGAACTCTTTGAGGAAGCACACATC 2811
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2812 TTTGGAATTACATTTTCTCCACCAGCAATGATGATTTTCTCTATGCTGGATGTTTC 2871
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 2872 CGGTTGGAGCCTACTTGGTGGCACAATAACTCATGAGCTTGAGGATGTTCTGTTAGTA 2931
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 2932 TTTTCAGCTGTGCTTGTGGTCATGCCGTGGGCAAGTCAGTTCAATTGCTGCCGAC 2991
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 2992 TATGCCAAAGCAAAATATCAGCAGCCACATCATCATGATCATTTGAAAAACCCCTTG 3051
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3052 ATTGACAGCTACAGCGGAGGCGCTAATGCCGAACACATTTGGAAGGAAATGTCACATTT 3111
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3112 GGTGAAGTTGATTCAACTATCCACCAGCCGACATCCAGTGTCTCAGGGAGCTGAGC 3171
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3172 CTGGAGGTGAAGAGGCCAGACGCTGCTGCTGGGAGCAGTGGCTGTGGGAAGAGC 3231
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3232 ACAGTGTCCAGCTCTCTGGAGCGGTTCTAGACCCCTTGGCAGGAAAGTGTCTGTTGAT 3291
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3292 GGCAGAAATTAAGCGCAATGATGTTTCAGTGGCTCCGAGCACACCTGGGCATCGTGTC 3351
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3352 CAGGAGCCCATCTCTGTTGACTGCAGCATTCCTGAGACATTCGCTATGGAGACACAGC 3411
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3412 CGGTGGTGTACAGGAAGAGATCGTGAGGCGCAAGAGGAGGCCAACATACATGCTCTC 3471

QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3472 ATCGAGTCACTGCTTAATAATATAGCACTAAAGTAGGACAAAGAACTCAGCTCTCT 3531
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3532 GGTGGCCAGAAACAACGCAATGCCATAGCTCGTGCCCTGTGTAGACAGCCTCATATTTG 3591
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGlnLysValValGlnGluAla 1219
Db 3592 CTTTGTGATGAAGCCAGCTCAGCTCGGTATGATGCTCGTCCCTGTGTAGACAGCCTCATATTTG 3651
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3652 CTGGACAAACCCAGAGAGGCCGCACTGCAATGTGATGCTCACCCTGTCCACCATC 3711
QY 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3712 CAGAATGCAGACTTAATAGTGGTGTTCAGAATGGCAGAGTCAAGGAGCATGGCACGCAT 3771
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3772 CAGCAGCTGCTGGCACAGAAAGCATCTATTTTCAATGTCAGTGTCCAGGCTGGAACA 3831
QY 1280 LysArg 1281
Db 3832 AAGCGC 3837

RESULT 7

ABA94365
ID ABA94365 standard; DNA; 3860 BP.
XX ABA94365;
AC ABA94365;
XX
DT 26-MAR-2002 (first entry)
XX Human BCRP DNA related seq Id No. 1.
DE
XX Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;
KW cardiac; gene therapy; ds.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..3843
FT /*tag= a
XX
PN WO200192877-A2.
XX
PD 06-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US17459.
XX
PR 31-MAY-2000; 2000US-0584586.
PR 29-MAY-2001; 2001US-0866866.
XX
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX Sorrentino B, Schuetz J;
XX
XX WPI; 2002-114368/15.
DR P-PSDB; ABH07266.
XX
PT Identifying a stem cell, for treating e.g., muscular dystrophy,
PT myocardial infarction, Parkinson's disease, or neurodegenerative
PT disorders, comprises detecting the expression of an ATP transport
PT protein (BCRP) by a cell
XX
PS Disclosure; Page 53-55; 87pp; English.
XX
CC The invention provides a method of identifying and/or isolating a stem

cell that involves detecting the expression of an ATP transport protein containing a conserved ATP-binding cassette (BCRP) by a cell in a sample comprising stem cells. The isolated stem cells may be used in the treatment of diseases such as muscular dystrophy, degenerative liver disorder, myocardial infarction, Parkinson's disease, degenerative CC disorders of the brain, and for tissue regeneration or replacement. CC Haematopoietic cells can be used in bone marrow transplants (e.g., for CC treatment of leukemia) and for ex vivo gene therapy for treating blood CC diseases such as sickle cell anemia and thalassemia. The stem cells can also be used as cell targets in gene therapy protocols. The present CC sequence represents a sequence related to the BCRP for which no relevant CC information has been provided in the specification.

xx
SQ Sequence 3860 BP; 1135 A; 746 C; 957 G; 1022 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	3860
Score:	5866.00	Matches:	1164
Percent Similarity:	95.09%	Conservative:	55
Best Local Similarity:	90.80%	Mismatches:	59
Query Match:	90.62%	Indels:	4
DB:	24	Gaps:	4

US-09-672-725c-25 (1-1281) x ABA94365 (1-3860)

Qy	1	MetAspProGluGlyArgLysGlySerAla---	GluLysAsnPheTrpLysMetGly	19
Db	1	ATGATCTTGAAGGGGCCCATGAGGAGGACAAAGAGAACTTTTAACTGAAC	60	
Qy	20	LysLysSerLysLysGluLysGluLysLysGluLysProThrValSerThrPheAlaMet	39	
Db	61	AATAAAAGT---GAAAAGATAAGAAAGAAACCAACTGTCAGTGTATTTTCAATG	117	
Qy	40	PheArgThrSerAsnTrpLeuAspArgLeuTrpMetLeuValGlyThrMetAlaAlaIle	59	
Db	118	TTTCGCTATTCAATGGCTTGACAGTGTATGATGGTGGGAACTTTGGCGGCATC	177	
Qy	60	IleHisGlyAlaAlaProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe	79	
Db	178	ATCCATGGGCTGGACTCTCTCATGATGCTGCTGTTGGAGAAATGACAGATATCTTT	237	
Qy	80	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr	99	
Db	238	GCAATATCAGGA---AATTAGAGATCTGATGTCAAAACATCACTAATAGAGTATATC	294	
Qy	100	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTrpAlaTrpTyr	119	
Db	295	AATGATACAGGTTCTTCATGAAT---CTGGAGGAGACATGACCATGCTATATAT	351	
Qy	120	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTrpIleGlnValSerPheTrpCys	139	
Db	352	TACAGTGGAAATGGTGTGGGGTCTGCTGCTGCTTACATTCAGGTTTCATTTTGGTGC	411	
Qy	140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159	
Db	412	CTGCAGCTGGAAGACAAATACACAAATTTAGAAAACAGTTTTTTCATGCTAATAATGCGA	471	
Qy	160	GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179	
Db	472	CAGAGATAGCTGGTTTGATGTGCACGATGTTGGGAGGCTTAACACCCGACTTACAGAT	531	
Qy	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle	199	
Db	532	GATGCTCTAAGATTAAAGGATTTGGTGACAAATTTGGAATTTGCTTTCAGTCAATG	591	
Qy	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuVal	219	
Db	592	GCAACATTTTTCACGTGGGTTTATAGTAGGATTTACAGCTGGTGGAGCTTAACCCCTGTG	651	
Qy	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239	
Db	652	ATTTTGGCCATCAGTCTCTTCTTGGACTGTCTGCTGCTGCTGGGCAAGATACTATCT	711	

Qy	240	SerPheThrAspLysGluLeuLeuAlaTrpAlaLysAlaGlyAlaValAlaGluGluVal	259	
Db	712	TCATTTACTGATAAAGAACTCTTAGCGGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTC	771	
Qy	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr	279	
Db	772	TTGGCAGCAATTAAGAACTGTGATTCATTTGGAGACAAAGAAAGAACTTGAAGGTAC	831	
Qy	280	AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299	
Db	832	AACAAAATTTAGAGAGACTTAAAGAAATGGGATAAAGAAAGCTATTACAGCCAATATT	891	
Qy	300	SerIleGlyAlaAlaPheLeuLeuIleTrpAlaSerTrpAlaLeuAlaPheTrpTyrGly	319	
Db	892	TCTATAGGTGCTGCTTTCTCTGATCTATGATCTTATGCTGCTGGCTTCTGTATGGG	951	
Qy	320	ThrSerLeuValLeuSerSerGluTrpSerIleGlyGlnValLeuThrValPhePheSer	339	
Db	952	ACCACCTTGGTCTCTCAGGGGAATATTCTATTGGACAAAGTACCTGATATTCTTTCT	1011	
Qy	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGlyAlaPheAlaAsn	359	
Db	1012	GTATTAATTTGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAT	1071	
Qy	360	AlaArgGlyAlaAlaTrpGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379	
Db	1072	GCAAGAGGAGCAGCTTATGAATCTTCAAGATAATTTGATAATAAGCCCAAGTATTGACAGC	1131	
Qy	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399	
Db	1132	TATTCGAAGATGGGCACAAACCATTAATATTAAAGGAAATTTTGAATTCAGAAATGTT	1191	
Qy	400	HisPheSerTrpProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419	
Db	1192	CACCTTCAGTTACCCATCTCGAAAAGAACTTAAGATCTTGAAGGGCTGAACCTGAAGGTG	1251	
Qy	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal	439	
Db	1252	CAGATGGGACAGCGTGGCTGTTGGAAACAGTGGCTGTGGGAGAGGACACACAGTC	1311	
Qy	440	GlnLeuMetGlnArgLeuTrpAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459	
Db	1312	CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGGATGGTCAGTGTGATGGACAGGAT	1371	
Qy	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro	479	
Db	1372	ATTAGGACCATAAATGTAAGGTTTCTACGGGAAATCATTTGGTGGTCAGTCAGGAACCT	1431	
Qy	480	ValLeuPheAlaThrThrIleAlaGluAsnIleArgTrpGlyArgGluAsnValThrMet	499	
Db	1432	GTATTTGTTGCCACCCAGCATAGCTGAAACATTCCTATGGCCGTGAAAATGTCAACATG	1491	
Qy	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaTrpAspPheIleMetLysLeuPro	519	
Db	1492	GATGAGATTGAGAAAGCTCTCAAGGAGCCCAATGCCTATGACTTTATCATGAACCTGCCT	1551	
Qy	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539	
Db	1552	CATAAATTTGACACCTGGTTGGAGAGAGGGGGCCAGTTGAGTGGTGGCGAGAGAGAG	1611	
Qy	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559	
Db	1612	AGGATCGCATTTGACAGTGGCTGGTTCGCAACCCCAAGATCTCTCTGCTGGATGAGGCC	1671	
Qy	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579	
Db	1672	ACGTACGCTTGGACACAGAAAGCCAGCAGTGGTTTCAAGTGGCTCTGGATAAGGCCAGA	1731	
Qy	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599	
Db	1732	AAAGGTGGGACCCACCATTTGTGATAGTCATCTGTTGCTACAGTTGCTGAATGCTGACGTC	1791	
Qy	600	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619	

Db 1792 ATFCGTGTTTCGATGATGGAGTCATTTGGAGAAAGGAAATCATGATGAATCATGAA 1851
Qy |||||||
Db 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Qy |||||||
Db 1852 GAGAAAGCATTTACTTCAAACTTGTCACAATGCAGACAGCAGGAATGAAGTTGAATTA 1911
Qy |||||||
Db 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Qy |||||||
Db 1912 GAAATGCAGCTGATGATCCAAAGTGAAATTGATGCTTGGAATGCTCTCAATGAT 1971
Qy |||||||
Qy 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Qy |||||||
Db 1972 TCAAGATCCACTTAATAAGAAAGATCAACTCGTAGGAGTGCCTGGATCAAGACC 2031
Qy |||||||
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGlnAsnValProProValSerPhe 699
Qy |||||||
Db 2032 CAAGCACAAAGCTTAGTACCAGAGGCTCTGGATGAAGATGATACCTCCAGTTTCCTTT 2091
Qy |||||||
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Qy |||||||
Db 2092 TGGAGATTATGAAGCTAAATTAACGAAATGGCTTATTTTGTGTTGGTGATTGTTGT 2151
Qy |||||||
Qy 720 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Qy |||||||
Db 2152 GCCATTATAAATGAGGCGCTGCAACCCAGCATTTGCAATAATATTTTCAAGATTTATAGG 2211
Qy |||||||
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Qy |||||||
Db 2212 GTTTTACAGAATGATGATCCCTGAAACAAACGACAGATAGTAATCTGTTTTCACATA 2271
Qy |||||||
Db 2272 TTGTTTCTAGCCCTTGGAAATTTATTTTATTACATTTTTCCTTCAAGGTTTCACATTT 2331
Qy |||||||
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Qy |||||||
Db 2332 GGCAAGCTGGAGAGATCCTCACCAAGCGCTCGATACATGTTTTCGATCCATCCTC 2391
Qy |||||||
Qy 800 ArgGlnAspValSerTyrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Qy |||||||
Db 2392 AGACAGATGTGAGTTGGTTGATGACCCCTAAACACCACTGGAGCATTTGACTACCAGG 2451
Qy |||||||
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Qy |||||||
Db 2452 CTCGCAATGATGCTGCTCAAGTTAAAGGGCTATAGSTTCCAGGCTTGCTGTAATACC 2511
Qy |||||||
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Qy |||||||
Db 2512 CAGAAATATAGCAAAATCTTGGACAGGAATAATTATATCTTCATCTATGGTTGGCAACTA 2571
Qy |||||||
Qy 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Qy |||||||
Db 2572 ACATGTTACTCTTAGCAATGTACCCTATTGCAATAGCAGAGTTGTTGAAATGAAA 2631
Qy |||||||
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 899
Qy |||||||
Db 2632 ATGTTGTCTGGCAAGCACGTGAAGATAAGAAAGAACTAGAGGTGCTGGGAAGATCGCT 2691
Qy |||||||
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Qy |||||||
Db 2692 ACTGAAGCAATAGAAAATCTCCGAACCGTTGTTTCTTCTGACTCAGGAGCAGAAAGTTTGA 2751
Qy |||||||
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Qy |||||||
Db 2752 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGAAAGCACACATC 2811
Qy |||||||
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Qy |||||||
Db 2812 TTTGGAATTACATTTTCTTCACCCAGCAATGATGATTTTTCCTATGCTGGATGTTTC 2871
Qy |||||||
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Qy |||||||

Db 2872 CGGTTTGGAGCCACTACTGGTGGCACATAAACTCATGAGCTTTTGAGGATGTTCTGTAGTA 2931
Qy |||||||
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Qy |||||||
Db 2932 TTTTACGCTGTTGCTTTTGTGTCATGGCGGTGGCAAGTCAGTTCATTTGCTCTCTGAC 2991
Qy |||||||
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Qy |||||||
Db 2992 TATGCCAAAGCCAAATATATCAGCAGCCACATCATCATGATCATTTGAAATAACCCCTTG 3051
Qy |||||||
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Qy |||||||
Db 3052 ATTCACAGCTTACAGCAGCGGAAGCCCTAAATGCCAGACACATTTGGAAGAAATGTCACATTT 3111
Qy |||||||
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Qy |||||||
Db 3112 GGTGAAGTTGATTTCAACTATCCACCCGACCGGACATCCAGTGTCTCAGGACATGAGC 3171
Qy |||||||
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Qy |||||||
Db 3172 CTGGAGGTGAAGAAGGCCAGACGCTGGCTGTGTGGCAGCAGTGGCTGTGGAGAGAC 3231
Qy |||||||
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuLeuAsp 1099
Qy |||||||
Db 3232 ACAGTGTCTCAGCTCCTGGAGCGGTTCTACGACCCCTTGGCAGGAAGTGTGCTTGAT 3291
Qy |||||||
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Qy |||||||
Db 3292 GGCAAGAAATAAAGCGACTGAATGTTCAGTGGCTCCGAGCACACCTGGGCTATCGTGCC 3351
Qy |||||||
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Qy |||||||
Db 3352 CAGGAGCCCATCTGTTGACTGCACCAITTCGTGAGAACATTCCTATATGAGACACAGC 3411
Qy |||||||
Qy 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
Qy |||||||
Db 3412 CGGGTGTGTACAGGAAGAGATCGTGGAGGCGCAAGAGGAGGCCAATACATGCTTTC 3471
Qy |||||||
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Qy |||||||
Db 3472 ATCGAGTCACTGCCTTAATAATATAGCACTAAGTAGGAGACAAAGAACTCAGCTCTCT 3531
Qy |||||||
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Qy |||||||
Db 3532 GGTGGCCAGAAACACACGATTCGCTAGCTCGTGCCCTGTTAGACAGCTTCATATTTG 3591
Qy |||||||
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Qy |||||||
Db 3592 CTTTGTGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAGCC 3651
Qy |||||||
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Qy |||||||
Db 3652 CTGGCAAAAGCCAGAGAGCGCCACCTGCATTTGATTTGCTACCGCTCTGCCACATC 3711
Qy |||||||
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Qy |||||||
Db 3712 CAGAATGCAGACTTAATAGTGTGTTTTCAGATGGCAGAGTCAAGGAGCATGGCAGCAT 3771
Qy |||||||
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Qy |||||||
Db 3772 CAGCAGCTGCTGGCAGAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 3831
Qy |||||||
Qy 1280 LysArg 1281
Qy |||||||
Db 3832 AGCGC 3837

RESULT 8

AAF86127

ID AAF86127 standard; cDNA; 4186 BP.

XX AAF86127;

XX

XX 25-JUN-2001 (first entry)

```
XX Cynomologous monkey P-glycoprotein cDNA.
DE
XX Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL;
KW efflux pump; ss.
XX
OS Macaca fascicularis.
XX
FH Key Location/Qualifiers
FT CDS 100..3942
FT /tag="a
FT /product="pgp"
FT /note="P-glycoprotein"
XX
XX W0200123565-A1.
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26592.
XX
XX 28-SEP-1999; 99US-0156921.
PR 12-OCT-1999; 99US-0158818.
XX
XX (GENT-) GENTEST CORP.
XX
XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
PI WPI; 2001-316136/33.
DR P-PSDB; AAB81064.
XX
XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell.
XX
XX Example 1; Page 51-57; 84pp; English.
XX
XX This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDRL is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents cDNA encoding
CC cynomologous monkey P-glycoprotein.
XX
XX Sequence 4186 BP; 1226 A; 799 C; 1039 G; 1122 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0 Length: 4186
Score: 5862.00 Matches: 1163
Percent Similarity: 95.32% Conservative: 59
Best Local Similarity: 90.72% Mismatches: 56
Query Match: 90.56% Indels: 4
DB: 22 Gaps: 4
US-09-672-725C-25 (1-1281) x AAF86127 (1-4186)
QY 1 MetAspProGluGlyArgLysGlySerAlaGlu---LysAsnPheTrpLysMetGly 19
Db 100 ATGGATCTTGAAGGGACCCCAATGAGGAGCAGAGAGAACTTTTAACTGAAC 159
QY 20 LysLysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 160 AATAAAGT---AAAAAAGATAGAGAGAGAGAAACCAACTGTCAGTGTATTTCAATG 216
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 217 TTTCGGTATTCAAAATGGCTTGACAAGTTGTATATGGTGGGAACACTTGGCTGCCATC 276
```

```
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 277 ATCCATGGAGCTGACCTCTCTCATGATGCTGTGTGTTGGAGACATGACGGATACCTTT 336
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 337 GCAAAATGCAGG---AATTTAGGAGATTAGGAGCTCTGTGTACTAATACAGCAATATC 393
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 394 ACTGATACAGTGCCTCATGAAT---CTGGAGGAAGATATGACAGGATGCTTATTAT 450
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 451 TACAGTGGAAATGGTGTGGGTGCTGCTTACATTACAGTTTCAATTTGGTGC 510
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 511 CTGGCAGCTGGAGACAAATACACAAATTTAGAAAACAGTTTTTTCATGCTATATAGCGA 570
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 571 CAGGATAGCTGGTTGATGTCACGATGTTGGGAGCTTAACACCGGCTTACAGAT 630
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 631 GATGCTCCAAGATTAATGAAGAAATGGTCACAAAATTTGGAATGTTCTTTCAGTCAATG 690
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 691 GCAACATTTTTCACCTGGGTTATAGTAGGATTTACAGCTGTTGGAAAGCTAACCTTGTG 750
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 751 ATTTTGGCCATCAGTCTCTGTTGGACTGTCAGCTGCGAGCTGGGCAAGATACTGTCT 810
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
Db 811 TCATTTACTATAAAGAACTCTTAGCTTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTC 870
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluAtrgTyr 279
Db 871 TTGCAGCAATTAAGACTGTGATTGCAATTTGGAGCACAAGAAAGAACTCGAAAGGTAC 930
QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 931 AACAAAAATTTAGAAAGACTTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCAATAT 990
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 991 TCTATAGGTGCTGCTTCTCTCTTATCTATGCACTTATGCTGTGGCTTCTGTGTATGGG 1050
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 1051 ACCACTTGGTCCTCAAGGAATATCTATTGACACAGTACTCCTGATTTCTTTCT 1110
QY 340 ValLeuIleGlyAlaPheSerIleGlycylAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1111 GTATTAATTTGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTTGC 1170
QY 360 AlaArgGlyAlaAlaTyrGluIlePheIleLysIleIleAspAsnLysProSerIleAspSer 379
Db 1171 GCAAGAGAGCAGCTTTTGAATCTTCAAGATAATTGATAATAAGCCAGATTTTGACAGC 1230
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1231 TATTGGAAGCTGGCACAACAGATAATATTAAAGGAAATTTTGAATTCAGAAATGTT 1290
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1291 CACTTTCAGTTACCCATCTCGAAAAAGAAAGTTAAGATCTTGAAGGGCCTTGAACCTGAAGGTG 1350
```

QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
|||||
Db 1351 CAGAGTGGCAGACGGTGGCCCTGGTTGGAAACAGCGGCTGGGAAGACACACCGTC 1410
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
|||||
Db 1411 CAGCTGATGTCAGAGCGCTTATGAGCCACAGAGGCGATGGTCAGTGTGTGAGCAGGAT 1470
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
|||||
Db 1471 ATTAGGACCAATACGTAAGGTTTCTACGGGAATCATCGGTGGTGGAGTCAGGAACCT 1530
QY 480 ValLeuPheAlaThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
|||||
Db 1531 GTATTGTTGGCCACCAGCATAGCTGAAACATTCGCTATGGTCGTGAAGATGTCACCATG 1590
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
|||||
Db 1591 GATGAGATTGAGAAAGCTGTCAAGGAAGCCCAATGCGCTATGACTTTATCATCAAACTGCCT 1650
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
|||||
Db 1651 CAGAAATTGACACCCCTGGTTGGAGAGAGAGGGCCCGAGCTGAGTGGTGGCAGAGCAG 1710
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAla 559
|||||
Db 1711 AGGATCGCCATGTCAGCTGCGCTGGTTGCGCAACCCCAAGATCCTCGCTGCGAGAGGCC 1770
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
|||||
Db 1771 ACGTCAGCCTTGGACACAGAAAGTGAAGCAGTGGTTTCAGTGGCTCTGGATAAGGCAG 1830
QY 580 LysGlyArgThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
|||||
Db 1831 AAAGTGGGACCATGATGAGTCACTCATCGTTGCTACGGTTCGTAATGCCGACGCT 1890
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
|||||
Db 1891 ATCGCTGGTTTCGATGATGAGTCACTCATCGTTGCTACGGTTCGTAATGCCGACGCT 1950
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
|||||
Db 1951 GAGAAAGGCATTTACTCAAACTTGTCACATGCACAGCAGGAAATGAAATTTGAATTA 2010
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
|||||
Db 2011 GAAATGTCAGCTGATGAATCCAAAGTGAATGATACCTTGGAAATGCTTCACATGAT 2070
QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
|||||
Db 2071 TCAGGATCCAGTCTAATAAGAAAGATCCACTCGTAGGAGTGTCCGTGGATCAACAAGC 2130
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
|||||
Db 2131 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 2190
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
|||||
Db 2191 TGGAGGATTATGAAGCTAAATTAATACTGAGTGGCCTTATTTGTTGTTGTTGTTGTTGTT 2250
QY 720 AlaIleIleAsnGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
|||||
Db 2251 GCCATTATAATGGAGGCTCGCAACAGCATTTGCGAGTAATATTTTCAAGATTTATAGG 2310
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
|||||
Db 2311 ATTTTACAGAAATGATGATCCGGAACCAACACAGAAATAGTAACATGTTTTCACATA 2370
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
|||||
Db 2371 TTGTTTCTAGTCTGGAATGTTTCTTTTATTACATTTTTCCTTCAGGCGCTTCACATTT 2430
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
|||||

Db 2431 GGCAAAAGCTGGAGAGATCCCTCACCAAGCGGCTCCGATACATGTTTTCGGATCATGCTC 2490
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
|||||
Db 2491 AGACAGGATGTGAGCTGGTTTGATGACCTTAAACACACCACTGGAGCATTGACTACGAG 2550
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
|||||
Db 2551 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTCTCTATAATTACC 2610
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleIleThrGlnLeu 859
|||||
Db 2611 CAGATATAGCAATCTTGGCAGGAATAATATATCTTAACTATCTATGTTGGCAACTG 2670
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
|||||
Db 2671 ACACCTGTTACTCTTAGCAATTGTACCCATCATTGCAATAGCAGGAGTCTTCAAAATGAA 2730
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
|||||
Db 2731 ATGTTCTCTGGACAAAGCACTGAAAGATAAGAAAGCACTAGAAAGGCTGGGAAGATCGCT 2790
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
|||||
Db 2791 ACTGAAGCAATAGAAACACTTCCGAACCTGTGTTTCTTTGACTCAGAGCAGAAAGTTTGA 2850
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
|||||
Db 2851 CATATGATGATCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATC 2910
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 2911 TTTGGAATCACCTTTTCTTCCACGAGGCAATGATGATTTTCTATCTGCTGGATGTTTC 2970
QY 960 ArgPheGlyAlaTyrLeuValAlaAlaHisValIleMetIleGluLysValLeuVal 979
|||||
Db 2971 CGTTTGGAGCCTACTTGGTGGCACATAGTCTCAGAGCTTTGAGGATGTTCTGTAGTA 3030
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
|||||
Db 3031 TTTTCAGCTGTTGTTGTTGGTGGCCATGGCGGTGGGCAAGTCAGTTTCATTTGCTCCTGAC 3090
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
|||||
Db 3091 TATGCCAAAGCCAAAGTATCAGACGCCCATCATCATGATCATGATGATGATGATGATGATG 3150
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
|||||
Db 3151 ATTGACAGCTACAGCACAGAGGCGCTAAAGCCGAACACATTTGGAAGGAAATGTCACATTT 3210
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
|||||
Db 3211 AATGAAGTTGATTAACAATATCCACCGAGCTGGACATCCAGCTGCTCAGGGGCTGAGC 3270
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
|||||
Db 3271 CTGGAAGTGAAGAGGGCCAGACGCTGGCCCTGTTGGCAGCAGTGGCTGTGGGAAGAGC 3330
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
|||||
Db 3331 ACGGTGGTCCAGCTCCTGGAGCGGTTCTATGACCCCTTGGCGGGGAAAGTCTGCTGTTGAC 3390
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
|||||
Db 3391 GGCAAAAGAAATAAGCAACTGAATGTTTCAGTGGCTCCGAGCACACCTGGGCATCGTCTCC 3450
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
|||||
Db 3451 CAGAGGCCATCCTGTTTGTGCTGAGCATTAGTGAGAACATTCCTATGAGGACAAAGCAGC 3510
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPhe 1159
|||||

Db 3511 CGGTGGTGTCTCAGAGAGATCGTGAGGCGAGCAAGGAGGCAATATACAGCCCTTC 3570
 QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
 Db 3571 ATCGAGTCACTCCCTAATAATATATAGACAGAGTAGGAGCAAAAGGAACCTCAGCTCTCT 3630
 QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
 Db 3631 GGTGGCGAAGAACAGCATTCGATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3690
 QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
 Db 3691 CTTTGGATGAAGCCACATCAGCTCTGGATACAGAAAGTGAAAGGTTGTCTCAAGAAGCC 3750
 QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
 Db 3751 CTGGCAAAAGCCAGAGAAAGCGGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3810
 QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
 Db 3811 CAGATGCGAGCTTAATAGTGTGTTTCAGATGCGAGAGTCAAGGAGCGACACAT 3870
 QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
 Db 3871 CAGCAGCTGCTGGCAGAGAAAGGCATCTATTTTCAATGGTCAGTGTCTCCAGGCTGGAGCA 3930
 QY 1280 LysArg 1281
 Db 3931 AAGCGC 3936

RESULT 9
 AAF86128
 ID AAF86128 standard; cdna; 4195 BP.
 AC AAF86128;
 XX 25-JUN-2001 (first entry)
 DE Cynomologous monkey P-glycoprotein cDNA variant 1.
 XX Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;
 KW eflux pump; ss.
 XX Macaca fascicularis.
 PH Key Location/Qualifiers
 FT CDS 100..3951
 FT /*tag= a
 FT /product= "pgp"
 FT /note= "P-glycoprotein"
 FT misc_feature 376..384
 FT /*tag= b
 FT /note= "Insertion of 9 nucleotides relative to PGP
 FT allelic variant AAF86127"
 XX WO200123565-A1.
 PN 05-APR-2001.
 XX 28-SEP-2000; 2000WO-US26592.
 PF 28-SEP-1999; 99US-0156921.
 PR 12-OCT-1999; 99US-0158818.
 XX (GENT-) GENTEST CORP.
 XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
 PI WPI; 2001-316136/33.
 DR P-PSDB; AAB81065.
 XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
 PT (PGP) and homologous PGP polypeptides are useful for predicting

bioavailability of compound and increasing PGP transporter activity in cell
 Example 1; Page 59-65; 84pp; English.
 This invention relates to a polynucleotide sequence encoding a cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the PGP protein. PGP, also known as multidrug transporter, MDR1 is a member of the ABC transporter superfamily. The enzyme serves as an efflux pump exporting small molecules across the cell membrane. The invention includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence and protein, and also that of an allelic variant. The PGP polynucleotide sequence is useful for increasing PGP transporter activity in a cell. Antisense sequences of the cDNA are useful for inhibiting PGP transport activity in a mammalian cell. They may also be used for increasing the bioavailability of a drug. The present sequence represents a cDNA encoding cynomologous monkey P-glycoprotein. This sequence contains a 9 nucleotide insert compared to the PGP allelic variant given in AAF86127.

Sequence 4195 BP; 1230 A; 801 C; 1039 G; 1125 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 4195
 Score: 5859.50 Matches: 1165
 Percent Similarity: 94.80% Conservative: 57
 Best Local Similarity: 90.38% Mismatches: 52
 Query Match: 90.52% Indels: 15
 DB: 22 Gaps: 5

US-09-672-725C-25 (1-1281) x AAF86128 (1-4195)

QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGlu---LysAsnPheTrpLysMetGly 19
 Db 100 ATGGATCTTGAAGGGGACCCGAATGGAGGAGCAGAGAAAGAACTTTTAACTGAAC 159
 QY 20 LysLysSerLysLysGluLysLysLysLysProThrValSerThrPheAlaMet 39
 Db 160 AATAAAAGT---AAAAAGATAAGAAAGAAAGAAACCACTGTCAGTGTATTTTCAATG 216
 QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
 Db 217 TTTCGCTATTCAAATGGCTTGACAAAGTTGTATATGGTGGGGAACCTTGGCTGCCATC 276
 QY 60 IleHisGlyValAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
 Db 277 ATCCATGAGCTGGACCTTCCTCTCATGATGCTGGTGTGGGAGACATCAGCATACCTTT 336
 QY 80 AlaAsnAlaGly-----IleSerArgAsnLysThrPheProVal 92
 Db 337 GCAATGCGAGAAATTTAGGAGATTTAGGAGCTCTGTTGTTTAAACAACACT----- 387
 QY 93 IleIleAsnGluSerIleThrAsnAsnThrGlnHisPheIleAsnHisLeuGluGluGlu 112
 Db 388 -----AATAGCAGTAATATCATCATACAGTCCCGCTCATGAAAT---CTGGAGGAAGAT 438
 QY 113 MetThrThrTyrAlaTyrTyrTyrSerGlyIleGlyAlaGlyValLeuValAlaIleTyr 132
 Db 439 ATGACACAGGTATGCCATATATTACAGTGGATTTGCTGGGGTCTGCTGCTGCTTAC 498
 QY 133 IleGlnValSerPheTrpCysLeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGln 152
 Db 499 ATTCAGGTTTCATTTTGGTGCCTGGCAGTGAAGAACAAATACACAAATATAGAAACAG 558
 QY 153 PhePheHisAlaIleMetArgGlnGluIleGlyTyrPheAspValHisAspValGlyGlu 172
 Db 559 TTTTTCATGCTATAATGCGACAGGAGATAGGCTGGTTGTATGTCACGATGCTGGGAG 618
 QY 173 LeuAsnThrArgLeuThrAspAspValSerLysIleAsnGluGlyIleGlyAspLysIle 192
 Db 619 CTTAACACCCGGCTTACAGATGATGCTCCAAGATAATGAAGGAATTTGGTACAAAAT 678
 QY 193 GlyMetPhePheGlnSerIleAlaThrPheThrGlyPheIleValGlyPheThrArg 212

Db	679	GGAAATGTTCTTTCAGTCAATGGCAACATTTTTCACGTGGGTTTATAGTAGGATTTACACGT	730
Qy	213	GlyTrpLysLeuThrLeuValIleLeuAlaIleSerProValLeuGlyLeuSerAlaAla	232
Db	739	GGTTGGAAAGCTAAACCCCTTGATTTTGGCCATCAGTCCTGTCTTGGACTGTCAGCTGCA	798
Qy	233	IleTrpAlaLysIleLeuSerSerPheThrAspLysGluLeuLeuAlaTyrAlaLysAla	252
Db	799	GTCGTGGCAAGATACTGCTCTTCATTACTGATAAAGAACTCTTAGCTTATGCAAAAGCT	858
Qy	253	GlyAlaValAlaGluValLeuAlaAlaIleArgThrValIleAlaPheGlyGlyCln	272
Db	859	GGACCACTAGCTGAGAGGCTCTGGCGCAANTTAGAAGCTGTGATTCATTTGGAGGACAA	918
Qy	273	LysLysGluLeuGluArgTyrAsnLysAsnLeuGluGluAlaLysGlyIleGlyLys	292
Db	919	AAGAAAGAACTCGAAAGGTACAACAAAAATTTAGAAGAAAGCTAAAAGAATTTGGGATAAG	978
Qy	293	LysAlaIleThrAlaAsnIleSerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyr	312
Db	979	AAAGCTATTACGCCAATATTTCTATAGGTGCTGCTTCTCTGCTTATCTATGCATCTTAT	1038
Qy	313	AlaLeuAlaPheTrpTyrGlyThrSerLeuValLeuSerSerGluTyrSerIleGlyCln	332
Db	1039	GCCTCGCCCTTCGTATGGGACCCTTGGTCCCTCAAGGAATATCTTATTGGACAA	1098
Qy	333	ValLeuThrValPhePheSerValLeuIleGlyAlaPheSerIleGlyClnAlaSerPro	352
Db	1099	GTACTCACTGATTCTTTCTGTATTAAATTGGGCTTTTAGTGTGGACAGCATCTCCA	1158
Qy	353	SerIleGluAlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheLysIleAlaEsp	372
Db	1159	AGCAATTGAAGCATTTGCCAATGCAAGAGAGCAGCTTTGAAATCTTCAAGATAATTGAT	1218
Qy	373	AsnLysProSerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGly	392
Db	1219	AATAAGCAAGATATTGACAGCTATTGGAAGGTGGGCAACACAGATAATATTAAAGGA	1278
Qy	393	AsnLeuGluPheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeu	412
Db	1279	AAATTTGGAATTCAGAAATGTTCACTTCACTTACCCTATCTCGAAAGAAGTTAAGATCTTG	1338
Qy	413	LysGlyLeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGly	432
Db	1339	AAGGGCTGAACTGAAGCTGACAGTGGGACAGCGTGGCCCTCGTGTGGAAACACGGCG	1398
Qy	433	CysGlyLysSerThrThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMet	452
Db	1399	TGTGGGAGAGACACACGGTCCAGCTGATGACAGAGCTTTATGACCCCAACAGAGGGCATG	1458
Qy	453	ValCysIleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThr	472
Db	1459	GTCAGTGTGTGACGACAGGATATTAGGACCATAACGTTAAGGTTTCTACGGGAATCATC	1518
Qy	473	GlyValValSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyr	492
Db	1519	GGTGTGGTCAGTCAGGAACCTCTATTGTTTGGCCACCAGATAGCTGAAACATTCGCTAT	1578
Qy	493	GlyArgGluAsnValThrMetAspGluIleGluLysAlaValLysGluAlaAsnAlaTyr	512
Db	1579	GGTGTGGAGATGTCAACATGATGAGATTGAGAAGCTGTTCAGGAAGGCCAATGGCTAT	1638
Qy	513	AspPheIleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGln	532
Db	1639	GACTTTATCATGAACTGCCTCAGAAATTTGCACCCCTGGTTGGAGAGAGGGGGCCAG	1698
Qy	533	LeuSerGlyGlyClnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLys	552
Db	1699	CTCAGTGGTGGGAGAGAAGCAGAGGATGCCATTGCAGTGGCCCTGGTTGCCAACCCCAAG	1758
Qy	553	IleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGln	572
Db	1759	ATCCTCTCTCGAGAGGCCAGCTCAGCCTTGGACACAGAAAGTGAAGCAGTGGTTCAG	1818

QY 933 SerLeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyr 952
 Db TCTTTGAGGAAGACACACATCTTTGGAATCAGGTTTCTTCACGCAGCAATGATGAT 2958
 QY 953 PheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsn 972
 Db TTTTCTATGCTGGATGTTTCGGTGTGGAGCTACTTGGTGCCACATAGTCTCATGAGC 3018
 QY 973 PheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGln 992
 Db TTTGAGGATGTTCTGTAGTATTTTTCAGCTGTGTCTTTGGTGCCATGGCCGTGGGGCAA 3078
 QY 993 ValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMet 1012
 Db GTGAGTTCATTTGCTCCTGACTATGCCAAGCCAAAGTATCAGCAGCCACATCATCATG 3138
 QY 1013 IleIleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThr 1032
 Db ATCATTTGAAAAAACCCCTTTGATTGACAGCTACAGCAGAGAAGGCTTAAGCCGGAACACA 3198
 QY 1033 LeuGluLysAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIle 1052
 Db TTTGGAAGGAATGTACATTTATGAAGTTGTATTCACACTATCCACCCGACTGGACATC 3258
 QY 1053 ProValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrLeuAlaLeuValGly 1072
 Db CCAGTCTTTCAGGGGCTGAGCCTGGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 3318
 QY 1073 SerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluLysArgPheTyrAspProLeu 1092
 Db AGCAGTGGCTGTGGGAAGACGCGTGGTCCAGCTCCTGGAGCGGTCTTATGACCCCTTG 3378
 QY 1093 AlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArg 1112
 Db GCGGGAAAGTCTGCTGACGGCAAGAAATAAAGCAACTGAATGTTTCAGTGGCTCCGA 3438
 QY 1113 AlaHisLeuGlyTleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsn 1132
 Db GCACACCTGGCATCTGTCGCCAGGAGCCATCTGTTTGAAGTGGAGCATTTAGTGAGAAC 3498
 QY 1133 IleAlaTyrGlyAspAsnSerArgValValSerHisGluGluIleMetGlnAlaAlaLys 1152
 Db ATTGCTATGGAGAACACACGCGGGTGTCCAGGAAGAGATCTGAGGCGACGCCAAG 3558
 QY 1153 GluAlaAsnIleHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGly 1172
 Db GAGGCCAATATACACGCTTCATCGAGTCACTGCCTTAATAATATAGCACACAGATAGGA 3618
 QY 1173 AspGlyThrGlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeu 1192
 Db GACAAAGGAACCTCAGCTCTCTGGTGCCCAAGAACACGCAATTTGCCATAGCTCGTCCCTT 3678
 QY 1193 ValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer 1212
 Db GTTAGACAGCTCATATTTTTCCTTTGGATGAGCCACATCAGCTCTGGATACAGAAAGT 3738
 QY 1213 GluLysValValGlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIle 1232
 Db GAAAAGGTTTCCAAAGAGCCCTGGCAAGAGCCAGAGAGGCGGTACCTGCATTTGTGATT 3798
 QY 1233 AlaHisArgLeuSerThrIleGlnAsnAlaAspLeuIleValValPheGlnAsnGlyLys 1252
 Db GCTCAGCGGCTGTCCACCATCCAGATGAGAGCTTAATAGTGGTGTTCAGATGGCAGA 3858
 QY 1253 ValLysGluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMet 1272
 Db GTCAAGGAGCAGCGCACATATCAGAGCTGCTGGCACAGAAAGGCATCTATTTTCAATG 3918
 QY 1273 ValSerValGlnAlaGlyAlaLysArg 1281
 Db GTCAAGTGTCCAGGCTGGAGCAAGAGCGC 3945

RESULT 10

AAZ49333
 ID AAZ49333 standard; cDNA; 3860 BP.
 XX AAZ49333;
 XX AAZ49333;
 DT 14-MAR-2000 (first entry)
 XX Human G185V mutant multidrug resistance-1 (MDR-1) cDNA.
 XX Multidrug resistance; MDR-1; P-glycoprotein;
 KW transmembrane efflux pump; haematopoietic stem cell; transduction;
 KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
 KW gene therapy; gene replacement; genetic defect; thalassaemia;
 KW Gauchier's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
 XX cytokine; mutant; ds.
 XX Synthetic.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FT CDS 1..3843
 FT /*tag= a
 FT /product= "Human G185V mutant MDR-1 protein"
 XX W09961589-A2.
 XX 02-DEC-1999.
 XX 27-MAY-1999; 99WO-US11825.
 XX 28-MAY-1998; 98US-0086988.
 XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 XX Sorrentino B, Bunting K;
 XX WPI: 2000-072615/06.
 XX P-PSDB: AAY58187.
 XX Ex vivo expansion of hematopoietic stem cells transduced with a
 sequence encoding human multidrug resistance-1, used for bone marrow
 transplantation -
 XX Example 1; Page 79-82; 113pp; English.
 XX This sequence represents cDNA encoding human G185V mutant multidrug
 resistance protein MDR-1, where the Gly residue at position 185
 of the wild-type protein (AAY58186) is replaced by Val. MDR-1 is a
 transmembrane efflux pump, responsible for the export of drugs from
 cells, particularly cancer cells. The wild-type MDR-1 shows increased
 resistance to ecoposide and decreased resistance to vinca alkaloids
 compared with the G185V mutant. The invention relates to transducing
 haematopoietic stem cells with nucleic acid encoding an MDR protein
 and culturing the modified cells. The modified haematopoietic stem
 cells are useful in bone marrow transplantation (to reconstitute
 haematopoietic systems in patients who have undergone chemotherapy or
 radiation therapy) and in ex vivo gene therapy of genetic defects in
 cells derived from haematopoietic stem cells, e.g., thalassaemia,
 Gauchier's disease, sickle cell anaemia or leukaemia. The modified
 cells can also be used to identify factors involved in regulating
 proliferation and differentiation in haematopoietic stem cells.
 XX Haematopoietic stem cells that express MDR-1 will be protected against
 chemotherapeutic agents, so can be engrafted while the patient is
 undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
 provides sufficient cells to permit standard biochemical analysis.
 XX Overexpression of MDR-1 allows cytokine-driven expansion of
 haematopoietic stem cells by at least 10-fold compared with a maximum
 of 4-fold in known procedures.
 XX Sequence 3860 BP; 1135 A; 746 C; 956 G; 1023 T; 0 other;

Alignment Scores: 0 Length: 3860
 Pred. No.: 0

[illegible]

Db 2032 CAAGACAGAAAGCTTAGTACCAGAGCGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 2091
Qy 700 TtpAqIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2092 TGGAGGATTATGAAGCTAAATTAACCTGAATGGCCCTTATTTGTTGTTGGTGATTTTGT 2151
Qy 720 AlaIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2152 GCCATTATAAGTGGAGCGCTGCAACAGCATTTGCAATAATATTTCAAAGATTATAGGG 2211
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2212 GTTTTTACAGAATTCATGATCCTGAAACAAAACGACAGATAGTAACCTGTTTTCACTA 2271
Qy 760 PhePheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2272 TTGTTTCTAGCCCTTGGAAATATTCTCTTTTATTACATTTTCTTCAAGTTTCACATTT 2331
Qy 780 GlyLysAlaGlyIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2332 GGCAAAGCTGGAGAGATCCTCACCAGCGCTCCGATACATGGTTTCCGATCCATGCTC 2391
Qy 800 ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2392 AGACAGGATGTAGTTGGTTGTATGACCTTAAACACCCACCTGGAGCATTTGACTACCCAGG 2451
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2452 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTCTGTAATTACC 2511
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2512 CAGAATATAGCAAACTCTGGACAGGAATAATATATCTTCATCTATGTTGGCAACTA 2571
Qy 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2572 ACACCTGTACTCTTAGCAATTTGACCCATCTTGCATATGCAATAGCAGGAGTTGTAAGATGAAA 2631
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2632 ATGTGCTGCGAACAGCAGCTGAAAGATAAGAAAGAACTAGAAAGTCTGGGAAGTCGCT 2691
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2692 ACTGAAGCAATAGAAACCTCCGAACCGTTGTCTTCTACCTCAGGAGCAGAGAGTTTGAA 2751
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2752 CATATGATGCTCAGAGTTGACAGTTACCATACGATAAGAACTCTTTGAGGAAGCACACATC 2811
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2812 TTTTGGAAATTACATTTCTTCCACCCAGGCAATGATGTATTTTCCCTATGCTGATGTTTC 2871
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 2872 CGGTTTGGAGCCCTACTTGGTGGCACATAAAGCTCATGAGCTTTGAGGATGTTCTGTAGTA 2931
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 2932 TTTTTCAGCTGTGTCTTGTGTCATGGCGTGGGCAAGTCAGTTCAITTCCTCTCGAC 2991
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 2992 TATGCCAAAGCCAAAATATCAGCAGGCCACATCATCATGATCATTTGAAAAAACCCTTTG 3051
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3052 ATTGACAGCTACACGCGGAAGCCCTAAATGCCGAACACATTTGGAAGAAATGTCACATTT 3111
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3112 GGTGAAGTTGTATTCAACTATCCACCCGACCGGACATCCACAGTGTCTCAGGAGCTGAGC 3171

Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3172 CTGCAGCTGAAGAAGGCCAGACGCTGGCTCTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3231
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuLeuAsp 1099
Db 3232 ACAGTGTCTCCAGCTCCCTGGAGCGGTTCTAGACCCCTTGGCAGGGAAGTGTCTGTATGAT 3291
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3292 GGCAAAAGAAATAAGCCAGCTGAATGTTTCAGTGGCTCGAGCACACCTGGGCATCGTGCC 3351
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3352 CAGGAGCCCATCTCTGTTTACTGCAGCATTTGCTGAGAACATTTGCTATGGAGACAACAGC 3411
Qy 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3412 CGGTGTGTGTCCAGGAGAGATCGTGAGGCGCAAGGAGGCCAACATACATACATGCTTC 3471
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3472 ATCGAGTCACCTGCCTAATAATATAGCACATAAGTAGGAGACAAAGAACTCAGCTCTCT 3531
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3532 GGTGCCAGAAACAACGATTCGCATAGCTGTGCTGCTGTAGACAGCCTCATATTTTG 3591
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3592 CTTTGTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAGCC 3651
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3652 CTGGACAAAGCCAGAGAGCGCCACCTGCATTTGATTGCTCACCCTGTCCACCATC 3711
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3712 CAGAATGCAGACTTAATAGTGTGTTTCAGAAATGGCAGAGTCAAGGAGCATGGCACCAT 3771
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3772 CAGCAGCTGCTGGCAGACAGAAAGGCATCTATTTTCAATGGTTCAGTCCAGGCTGGAACA 3831
Qy 1280 LysArg 1281
Db 3832 AAGCGC 3837
RESULT 11
ABA94366
ID ABA94366 standard; DNA; 3860 BP.
XX ABA94366;
XX
XX 26-MAR-2002 (first entry)
DT
XX
DE Human BCRP DNA related seq Id No. 3.
XX
XX Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
KW hepatotropic; neurodegenerative; cytosolic; antianemic; muscular; BCRP;
KW cardiant; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 1..3843
FT /*tag= a
XX
XX W0200192877-A2.
XX
XX 06-DEC-2001.
XX

Db	1492	GATGAGATTGAGAAAGCTGTCAAGGAAGCAATGCTATGACTTTATCATGAAACTGCCT	1551
Qy	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539
Db	1552	CATAAAATTTGACACCTGGTTGGAGAGAGGGGCCAGTTGAGTGGTGGCGAAGAAGCAG	1611
Qy	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuLeuAspGluAla	559
Db	1612	AGGATCGCCATTGCACGTCCTGGTTGCAACCCCAAGATCCTCCTCGCTGGATGAGGCC	1671
Qy	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579
Db	1672	ACGTGACGCTTGGACACAGAAAGCAAGCAGTGGTTCAAGTGGCTCGGATAAAGGCCAGA	1731
Qy	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599
Db	1732	AAAGGTGGACACCACCATTTGTGTAGCTCATCGTTTGTCTACAGTTCGTAAATGCTGAGTC	1791
Qy	600	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619
Db	1792	ATCCTCGTTTTCGATGATGAGTCATTGTGGAGAAGGAATCATGATGAACATCATGAAA	1851
Qy	620	GluLysGlyIleTyrrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu	639
Db	1852	GAGAAAGCATTTACTTCAAACTGTGTCAATGCAGACAGCAGGAAATGAAGTTGAATTA	1911
Qy	640	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	659
Db	1912	CAAAATTCACGCTGATGAATCCAAAAGTGAATGTGATGCTTGGAAATGTCTTCAAAATGAT	1971
Qy	660	SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly	679
Db	1972	TCAGATCCAGTCTTAATAGAAAAGATCAACTCGTAGAGTGCCTGGGTGCACAAGCC	2031
Qy	680	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	699
Db	2032	CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGTATATACCTCCAGTTTCTCTT	2091
Qy	700	TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrrPheValValGlyIlePheCys	719
Db	2092	TGGAGGATTAATGAAGCTAAATTTAACTGAATGGCCTTAATTTGTTGTTGGTATATTTGT	2151
Qy	720	AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	739
Db	2152	GCCATTAAATGAGGCTTCGAACGACGCTTTCGAATAATATTTTCAAGATTTATAGGG	2211
Qy	740	IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetSerVal	759
Db	2212	GTTTTTCAAGAATTGATGATCTCGAAACAAACACAGATAAGTAACTTGTGTTTCACTA	2271
Qy	760	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	779
Db	2272	TTGTTTCTAGCCCTTGGAAATATTTCTTTTATATACATTTTTCCTTCAAGGTTTCACATTT	2331
Qy	780	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrrMetValPheArgSerMetLeu	799
Db	2332	GGCAAGCTGGAGAGATCCTCACCAAGCGCTCCGATACATGGTTTCCCGATCCATGCTC	2391
Qy	800	ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg	819
Db	2392	AGACAGGATGTGAGTTGTTGTATGACCTAAAAACACCACTGGAGCATTCGACTACCAGG	2451
Qy	820	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	839
Db	2452	CTCGCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGCGCTTGCCTGAATTACC	2511
Qy	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrrGlyTrpGlnLeu	859
Db	2512	CAGAAATATAGCAATCTTGGACAGGAATAATATATCTTCATCTATGTTGGCAACTA	2571
Qy	860	ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	879

QY 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
 |||||
 Db 3712 CAGATGCAGACTTATAGTGGTCTTTCAGAAATGGCAGAGTCAAGAGCATGGCACGCAT 3771
 QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
 |||||
 Db 3772 CAGCAGCTGCTGGCAGAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 3831
 QY 1280 LysArg 1281
 |||||
 Db 3832 AAGCGC 3837

RESULT 12

AAH57442

ID AAH57442 standard; cDNA; 4349 BP.

XX

AC AAH57442;

XX

DT 10-SEP-2001 (first entry)

XX

DE Human intestine cell specific cDNA sequence SEQ ID NO:282.

XX

KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
 KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
 KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
 KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.

XX

OS Homo sapiens.

XX

PN WO200132927-A2.

XX

PD 10-MAY-2001.

XX

PF 02-NOV-2000; 2000WO-US30396.

XX

PR 04-NOV-1999; 99US-0163508.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Sornasse T, Sellhamer JJ, Watson GA;

XX

DR WPI; 2001-291057/30.

XX

PT New cell and tissue specific polynucleotides useful for diagnosis,
 PT prognosis or monitoring of treatments for disorders where the gene is
 PT associated with a cancer; immunopathology or neuropathology -

XX

PS Claim 1; Page 207-208; 327pp; English.

XX

CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
 CC sequences (I). (I) can have cytostatic, immunomodulatory and
 CC neuroprotective activities, and can be used in gene therapy. (I) and
 CC proteins (II) encoded by then are used in high throughput screening
 CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,
 CC mimetics, peptides, proteins, agonists, antagonists, antibodies or
 CC their fragments, immunoglobulins, inhibitors, drug compounds and
 CC pharmaceutical agents. Expression of (I) in a sample indicates the
 CC differentiation of embryonic stem cells into a tissue selected from
 CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
 CC tissues. (I) and (II) are used to produce an expression profile that
 CC defines a metabolic or developmental process, treatment, condition,
 CC disease or disorder. The gene profile can be used for diagnosis,
 CC prognosis or monitoring of treatments and for investigating a
 CC predisposition to a disorder where the gene is associated with a
 CC cancer, immunopathology or neuropathology.

XX

SQ Sequence 4349 BP; 1304 A; 816 C; 1048 G; 1181 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 4349

Score: 5857.00 Matches: 1163

Percent Similarity: 95.01% Conservative: 55

Best Local Similarity: 90.72% Mismatches: 60
 Query Match: 90.48% Indels: 4
 DB: 22 Gaps: 4
 US-09-672-725c-25 (1-1281) x AAH57442 (1-4349)
 QY 1 MetAspProGluGlyGlyArgLysGlySerAla--GluLysAsnPheTrpLysMetGly 19
 |||||
 Db 126 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAAAGAACTTTTTTAAACGAAC 185
 |||||
 QY 20 LysLysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
 |||||
 Db 186 AATAAAAGT---GAAAAAGATAAGAGAGAAAGAAACCAACTGTCAGTGTATTTTCAATG 242
 |||||
 QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
 |||||
 Db 243 TTTTCGCTATTCAAAATTTGGCTTGACAAGTTGATATGTTGGGAAATGACAGATATCTTT 302
 |||||
 QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
 |||||
 Db 303 ATCCATGGGGCTGGACATTCCTCTCATGATGCTGTTGGGAAATGACAGATATCTTT 362
 |||||
 QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
 |||||
 Db 363 GCAAATGCAGGA---AATTTAGAAGATCTGATGTCAAACATCACTAATAGAGTGATATC 419
 |||||
 QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 119
 |||||
 Db 420 AATGATACAGGGTCTTTCATGAAT---CTGGAGAGACATGACCAGGTATGCCTATTAT 476
 |||||
 QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
 |||||
 Db 477 TACAGTGAATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 536
 |||||
 QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
 |||||
 Db 537 CTGCGAGCTGGAAGACAAATACACAAAATTAGAAAACAGTTTTTTCATGCTATAATGCGA 596
 |||||
 QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
 |||||
 Db 597 CAGGAGATAGGCTGGTTGATGTGCACGATTTGGGAGGCTTAACACCGGACTTACAGAT 656
 |||||
 QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
 |||||
 Db 657 GATGCTCTTAAGATTTAATGAAGTTATTTGGTACAAAATTTGGAATGTTCTTCACTCAATG 716
 |||||
 QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
 |||||
 Db 717 GCAACATTTTTCACCTGGGTTTATAGTAGGATTTTACACGTTGGTGGAGGCTAACCCCTTGG 776
 |||||
 QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
 |||||
 Db 777 ATTTTGGCCATCAGTCTCTTCTTGGACGTCTCAGCTGCTCTGGGCAAGAGTACTATCT 836
 |||||
 QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
 |||||
 Db 837 TCATTACTGATATAAGACTCTTAGCGTATGCAAAAGCTGGAGCAGTACTGAAGAGGTC 896
 |||||
 QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
 |||||
 Db 897 TTGGCAGCAATTAAGAAGCTGTGATTGCAATTTGGAGGACAAAAGAAAGAACTTGAAGGTAC 956
 |||||
 QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
 |||||
 Db 957 AACAAAAATTTAGAAAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCCAATATT 1016
 |||||
 QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
 |||||
 Db 1017 TCTATAGTGGCTGCTTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATG 1076
 |||||
 QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
 |||||
 Db 1077 ACCACCTTGGTCTCTCAGGGGAATATTCTATTGGACAGTAGTACTCAGTGTATTTCTTCT 1136
 |||||

Qy	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359
Db	1137	GTATTAAATTGGGGCTTTAGCTTGGCAGCGCATCTCCAGCATTTGAAGCATTTGC	1196
	360	AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379
Db	1197	GCAAGAGGAGCAGCTTATGAAATCTTCAAGATAATTGATATAAGCCAAGTATTGCAGC	1256
Qy	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399
Db	1257	TATTTCGAAGAGTGGGCACAAACAGATAATATTAGGGAAATTTGGAAATTCAGAAATGTT	1316
Qy	400	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419
Db	1317	CACCTTCAGTTACCATCTCGAAGAAAGATTAGATCTTTGAAGGCCCTGACCTGGAAGTGTG	1376
Qy	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal	439
Db	1377	CAGATGGGCAGACGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGACGACCAACAGTC	1436
Qy	440	GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459
Db	1437	CAGCTGATGCAGAGGCTCTATACGCCCCACAGAGGGAGTGGTCAGTGTGTGATGGACAGAT	1496
Qy	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro	479
Db	1497	ATTAGGACCATAAATCTAAGTTCTTACGGGAAATCATTTGGTGGTGAGTCAGGAACCT	1556
Qy	480	ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499
Db	1557	GTATTGTTTGCCACACAGATAGCTGAAACATCTCCCTATGCCGTGAAATGTCAACCATG	1616
Qy	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro	519
Db	1617	GATGAGATTGGAAGACTGTCAAGGAGCCATGGCTATGACTTTATCATGAAACTGCCT	1676
Qy	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539
Db	1677	CATAAATTTGCACCCCTGGTTGGAGAGAGAGGGGCCAGTTTCAGTGGTGGCGAGAAGCAG	1736
Qy	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559
Db	1737	AGGATCGCCATTTCAGCTGCCCTGGTTTGCACACCCACCAAGATCTCTCTGCTGGATGAGCC	1796
Qy	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579
Db	1797	ACGTCAGCCTTGGACACAGAAAGCAGAGCAGTGGTTTCAGTGGCTCTCGATAAGGCCAGA	1856
Qy	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599
Db	1857	AAAGTCGGACCACTTGTGTAGTCACTATCGTTTGTCTACAGTTTCGTAATTCGTGAGCTC	1916
Qy	600	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619
Db	1917	ATCCCTCGTTTCGATGATGGAGTCATTGTGGAAAGGAAATCATGATGAATCATGAA	1976
Qy	620	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu	639
Db	1977	GAGAAAGGCATTTACTTCAAACTTGTCAAAATGCAGACAGCAGGAGAAATGAAGTTGAATTA	2036
Qy	640	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	659
Db	2037	GAAATCGACGTGATGAATCCAAAAGTGAAATGTGATGGCTTGGAAATGCTTCAAAATGAT	2096
Qy	660	SerGlySerSerLeuLysArgArgSerThrArgArgSerIleHisAlaProGlnGly	679
Db	2097	TCAAGATCCAGTCTAATAAGAAAAAGATCAACTCGTAGGAGTGTCCGTGGATCATCAAGCC	2156
Qy	680	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	699
Db	2157	CAACAGAAAGCTTGTAGTACCAAGAGGCTCTGGATGAAAGTATACCTCCAGTTTCCTTT	2216

Qy	700	TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	719
Db	2217	TGGAGGATTATGAAGCTAAATTAACCTGAATGGCCTATTGTGTGGTGAATTTTGT	2276
Qy	720	AlaIleIleasnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	739
Db	2277	GCCATTATAAATGGAGCCTGCAACGAGCATTTGCAATAATATTTTCCAAAGATTATAGG	2336
Qy	740	IlePheThrArgAspGlnAspProGlnThrLysArgGlnAsnSerAsnMetPheSerVal	759
Db	2337	GTTTTTACAGAATTCATGATCCTGAACAAACACGACAGATAGTAACTGTGTTTCACAT	2396
Qy	760	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	779
Db	2397	TTGTTTCTAGCCCTTGGAAATTATTTCTTTATTACATTTTCCCTCAGGCTTTACATTT	2456
Qy	780	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	799
Db	2457	GGCAAAAGCTGGAGAGATCCTCCACCAAGCGCTCCGATACATGTTTCCGATCCATGCTC	2516
Qy	800	ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg	819
Db	2517	AGACGAGATGTGATGTTGGTTGATGCCCTAAAACACACCATCGAGCATTTGACTACCAG	2576
Qy	820	LeuAlaAsnAspAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	839
Db	2577	CTCGCCAATCATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTAATTACC	2636
Qy	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	859
Db	2637	CAGAATATAGCAAAATCTGGGACGAGAAATATATATATCTTCATCTATGTTGGCAACTA	2696
Qy	860	ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	879
Db	2697	ACATGTTTACTCTTAGCAATGTGACCCATCATTCGAATGACGAGTGTGTGAATGAAA	2756
Qy	880	MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla	899
Db	2757	ATGTTGTCTGGCAAGCACTGAAAGATAAGAAAGAACTAGAAAGGTGCTGGGAAGATCGCT	2816
Qy	900	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu	919
Db	2817	ACTGAAGCAATAGAAAACCTTCCGAACCGTGTCTTGTGACTCAGGAGCAGAGTTTGAA	2876
Qy	920	TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle	939
Db	2877	CATATGTATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTCAGGAAGACACATC	2936
Qy	940	PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe	959
Db	2937	TTTGGAAATACATTTTCTTCCACCCAGGCAATGATATTTTCTATGCTGGAATGTTTC	2996
Qy	960	ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValIleLeuVal	979
Db	2997	CGGTTTGGAGCCTACTTGGTGGACATAAATCATGAGCTTTGAGGATGTTCTCTTAGTA	3056
Qy	980	PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp	999
Db	3057	TTTTTCAGCTGTGTCTTGGTGCCATGGCGCTGGGGCAAGTCAGTTCATTTGCTCCTGAC	3116
Qy	1000	TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu	1019
Db	3117	TATGCCAAAGCCAAAATATCAGCAGCCACATCATCATGATCAATGAAAACCCCTTTG	3176
Qy	1020	IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe	1039
Db	3177	ATTGACAGCTACAGCACGGAAGCCCTAAATCCCGAACACATTTGGAAGGAAATGTCCACATTT	3236
Qy	1040	AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer	1059
Db	3237	GGTGAAAGTTGTATCAACTATCCACCCGACGACATCCAGTCAGTCTTTCAGGACTGAGC	3296
Qy	1060	LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer	1079

```
Db 3297 CTGAGGTGAAGAAAGGCGACAGCCTGGCTCTGGTGGCGACGAGTGGCTGGGAAGAGC 3356
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3357 ACATGGTCCAGCTCCCGGAGCGGTCTACGACCCCTGGCAGGGAAGAGTGGCTTGAT 3416
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3417 GGCAAAGAAATAAAGCGACTGAATGTTTCAGTGGCTCGAGCACACCTGGGCATCGTGCC 3476
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3477 CAGAGGCCCATCTCTGTTGACTGCAGCATTCCTGAGAACATTCGCTATGGAGACAACAGC 3536
Qy 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3537 CGGTGGTGTGCACAGGAAGAGATCGTGGGGCAGCAAGGAGGCCAACATACATAGCCTTC 3596
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3597 ATCCAGTCACTGCCTAATAATATAGCACTAAAGTAGGACAAAGAACTCAGCTCTCT 3656
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3657 GGTGGCCAGAAACAACGATTCGCATAGCTCGTCCCTTGTAGACAGCCTCATATTTG 3716
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3717 CTTTGTGATGAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGGTTGTCCAAGAGCC 3776
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3777 CTGGACAAAGCCAGAGAAAGCCGACCTGCATTGTGATTGCTCACCCTGTCACCATC 3836
Qy 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3837 CAGAATGCAGACTTAATAGTGGTGTTCAGAATGGCAGAGTCAAGGAGCATGGCAGCAT 3896
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3897 CAGCAGCTGTCGCACAGAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 3956
Qy 1280 LysArg 1281
Db 3957 AAGCGC 3962

RESULT 13
AAZ94738
ID AAZ94738 standard; cDNA; 4646 BP.
XX
AC
XX
XX
DT
XX
DE
XX
KW ABCB1; ATP binding cassette ABCB1 (MDR1) cDNA.
KW atherosclerosis; human; cholesterol; lipid disorder;
KW lupus erythematosus; lipid disorder; dyslipidemia; psoriasis;
KW multidrug resistance; diagnosis; gene therapy; MDR1;
XX
OS Homo sapiens.
XX
XX WO200018912-A2.
XX
XX 06-APR-2000.
XX
XX 21-SEP-1999; 99WO-EF06991.
XX
XX 25-SEP-1998; 98US-0101706.
XX
PA (FARB ) BAYER AG.
```

```
XX Schmitz G, Klucken J;
PI WPI; 2000-293151/25.
XX
DR Adenosine triphosphate binding proteins useful for identifying agents
XX for treating atherosclerosis and other inflammatory disorders -
XX
PS Claim 9; Page 110-112; 154pp; English.
XX
CC The present sequence is that of human ATP binding cassette
CC subfamily B protein ABCB1 cDNA. The cDNA was identified using a
CC differential display method in which monocytes from peripheral
CC blood were subjected to macrophage differentiation and cholesterol
CC loading with acetylated low density lipoproteins and subsequent
CC deloading with high density lipoprotein (HDL3) to identify
CC cholesterol sensitive genes. The gene maps to chromosome 7q21
CC and is also termed MDR1 (multidrug resistance). The invention
CC provides cholesterol-sensitive ABC genes (see AA294734-63). These
CC genes, and polypeptides encoded by them, can be used for diagnostic
CC and therapeutic applications, and for biochemical or cell-based
CC assays to screen for pharmacologically active modulator compounds
CC useful for the treatment of lipid disorders, atherosclerosis or
CC other inflammatory diseases such as psoriasis and lupus
CC erythematosus.
XX
SQ Sequence 4646 BP; 1371 A; 892 C; 1129 G; 1254 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 4646
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 21 Gaps: 4

US-09-672-725C-25 (1-1281) x AA294738 (1-4646)

Qy 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 425 ATGGATCTTTGAAGGGGACCGCAATGGAGGAGCAAGAAGAAAGAACTTTTAAACTGAAAC 484
Qy 20 LysLysSerLysLysLysLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 485 AATAAAAGT---GAAAGATGAAGAGGAAAGAAACCACTGTCAGTGTATTTTCAATG 541
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 542 TTTTCGCTATTCAAATTTGGCTTGACAAAGTTGTATATGTTGGTGGGAACCTTTGGCTGCCATC 601
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 602 ATCCATGGGCTGGACTTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 661
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 662 GCAATATGCAGGA---AATTAGAGATCTGATGTCACAAACATCACTAATAGAGTGCATATC 718
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 719 AATGATACAGGGGTTCTTCATGAAT---CTGGAGGAAGACATGACAGGATGCGCTATTAT 775
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 776 TACAGTGAATTTGGTGGGGTGTCTGCTTCTACATCAGGTTTCAATTTTGGTGC 835
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 836 CTGGCAGCTGGAAGACAAATACACAAATATAGAAACAGTTTTTTCATGCTATATAATCGGA 895
Qy 160 GlnGluIleGlyTyrPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 896 CAGGAGATAGGCTGTTTGTATGTCACGATGTTGGGAGCTTAACACCCGACTTACAGAT 955
```

QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
DB 956 GATGTCCTAAAGATTAAATGAAGTTATTTGGTGACAAAATGGAAATGTTCTTCAGTCAATG 1015
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
DB 1016 GCACATTTTTCACYGGGTTTATAGTAGGATTTACACGTGGTTGGAAGCTAACCCCTTGTG 1075
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleThrAlaLysIleLeuSer 239
DB 1076 ATTTTGGCCATCAGTCCTCTCTTGACGTGTCAGCTGCTGTGGCRAAGATACATATCT 1135
QY 240 SerPheThrAspLysGlnLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
DB 1136 TCATTTACTGATANAAGACTCTTAGCGTATGCANAAGCTGGAGCAGTAGCTGAAGAGGTC 1195
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
DB 1196 TTGGCAGCAATTAGAACTGTGATTGCTATTTGGAGGACAAAAGAACTTCAAAGGTAC 1255
QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
DB 1256 AACAAAAATTTAGAAGAAGCTAAAGAATTTGGGATAAAGAAAGCTATTACAGCCAATATT 1315
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
DB 1316 TCTATAGGTGCTCTTCTCGTGATCTATGCATCTTATGCTCTGGCCTTCTGGTATGG 1375
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
DB 1376 ACCACCTTGGTCTCTCAGGGGAATATCTATTGGACAAGTACTCAGCTATTTCTTTTCT 1435
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
DB 1436 GTATTAAATGGGGCTTTTAGTGTGGACAGGCATCTCCAAGCATTGAACATTTGCAAT 1495
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
DB 1496 GCAAGAGGACAGCTTATGAATCTTCAAGATAATTTGATAAAGCCAAGTATTGACAGC 1555
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
DB 1556 TATTCGAAGAGTGGGCACAAACCAGATAATATTAAAGGAAATTTGGAATTCAGAAATGT 1615
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
DB 1616 CACTTCAGTTACCCATCTCGAAAAGAAGTTAAGATCTTGAAGGCCCTGAACCTGAAGGTG 1675
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
DB 1676 CAGAGTGGGCAGACGGTGGCCCTTGGTTGGAACAGTGGCTGTGGGAAGAGCACAAACAGTC 1735
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
DB 1736 CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGATGTGAGTGTGATGACACAGGAT 1795
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
DB 1796 ATTAGGACCATAAATGTAAGGTTTACGGGAAATCATTTGGTGTGTGAGTCAGGAACCT 1855
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
DB 1856 GTATTGTTGCCACACGATAGCTGAACAACATTCGCTATGGCCGTGAAAATGTCACCATG 1915
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
DB 1916 GATGAGATTGAGAAAGCTGTCAGGAAGCCCAATGCCCTATGACTTTATCAAGAACTGCCT 1975
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
DB 1976 CATAAATTTGACACCCCTGGTTGGAGAGAGAGGGGCCAGTTGAGTGTGGCGGCAAGACGAG 2035

QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
DB 2036 AGGATCGCCATTGCACGTGCCCTGGTTGCGAACCCCAAGATCCTCCTGCTGGATGAGGCC 2095
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
DB 2096 ACGTACGCTTGGACACAGAAAGCGAAGCAGTGGTTCAGGTGGCTCGGATAAGGCCAGA 2155
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
DB 2156 AAAGCTGGACCCACCATTTGTATAGCTCATCGTTGCTACAGTCTGTAATGCTGACGTC 2215
QY 600 IleAlaGlyPheAspAspGlyValIleValGlnLysGlyAsnHisAspGluLeuMetLys 619
DB 2216 ATCGCTGGTTTCGATGATGGAGTCATTGTCGAGAAAGAAATCATGATGAACATGAAA 2275
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
DB 2276 GAGAAAGCATTTACTTCAAACTTGTCAATGCGACATGCGACAGCAGGAAATGAAGTTGAATTA 2335
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
DB 2336 CAAATGCAGCTGATGAATCCAAAAGTGAATTCATGCCCTTGGAAATGCTCTTCAAAATGAT 2395
QY 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
DB 2396 TCAAGATCCAGTCTAATAAGAAAAAGATCAACTCGTAGGAGTGTCCGTGGATCAACAAGCC 2455
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
DB 2456 CAACACAGAAAGCTTAGTACCNAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCCTT 2515
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
DB 2516 TGGAGGATTATGAAGCTAAATTTAACTGAATGGCTATTATTTGTTGTTGTTGTTATTTGT 2575
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
DB 2576 GCCATTATAAATGGAGGCTCGAACCCAGCATTTGCAATAATATTTTCAAGATTTATAGG 2635
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
DB 2636 GTTTTTCACAGAAATGTATGATCCTGAAACAAAACACACAGAAATAGTAACCTGTTTTCAC 2695
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
DB 2696 TTGTTTCTAGCCCTTGGAAATTTATCTTTTATACATTTTCTCAGGGTTTCCACATTT 2755
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
DB 2756 GGCAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATGGTTTCCGATCCATGCTC 2815
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
DB 2816 AGACAGGATGTAGTTGGTTTGTATGACCTAAAAACACACCTGGAGCATTTGACTACCAGG 2875
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
DB 2876 CTCGCCAATGATGTGCTCAAGTAAAGGGGCTATAGTTCAGGCTTCGCTGTAATTACC 2935
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
DB 2936 CAGATATAGCAAACTTTGGGACAGGAATAATATATCTCTCATCTATGTTGGCAACTA 2995
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
DB 2996 ACACCTGTTACTCTTAGCAATTTGTACCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAAA 3055
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
DB 3056 ATGTTGCTGGACAAGCACTGAAAGATAGAAGAACTAGAAGGTGCTGGGAGATCGCT 3115
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919

Db 3116 ACTGAAGCAATAGAAAACCTCCGAACCGTGTCTTCTTGTACAGGAGCAGCAACTTTGAA 3175
 QY TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
 Db 3176 CATATGTATGTCAGAGTTTCAGGTACCATACATACAGAACTCTTTGAGGAAAGCACATC 3235
 QY PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
 Db 3236 TTTGGAATTACATTTCTTCACCCAGGCAATGATATTTTCTATGCTGGATGTTT 3295
 QY ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
 Db 3296 CGGTTTGAGCGCTACTTGGTGCCACATAAATCATGAGCTTTGAGGATGTTCTGTAGTA 3355
 QY PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
 Db 3356 TTTTCAGCTGTGCTTTGGTGCCATGGCGTGGGCAAGTCACTGATTTGCTCTGAC 3415
 QY TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleLeuGluLysSerProLeu 1019
 Db 3416 TATGCCAAAGCCAAATATCAGCAGCCACATCATCATGATCATTTGAAAACCCCTTTG 3475
 QY IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
 Db 3476 ATTGACAGCTACAGCAGGAGGCGCTAATGCCGAACACATTTGGAAGGAATGTACATTT 3535
 QY AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
 Db 3536 GGTGAAGTTGATTCAACTATCCACCCGACCCGACATCCAGTGTCTTCAGGAGCTGAGC 3595
 QY LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
 Db 3596 CTGGAGGTGAAGAGGCCAGAGCGCTGGCTCTGGTGGGACAGTGGCTGTGGAGAGAGC 3655
 QY ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
 Db 3656 ACAGTGTCTCAGCTCTCTGGAGCGGTCTTACGACCCCTTGGCAGGAAAGTGTCTTGAT 3715
 QY GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
 Db 3716 GGCAGAAATAAAGCAGCTGAATGTTTCACTGGCTCCGAGCACACCTGGGCACTCGTTC 3775
 QY GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
 Db 3776 CAGAGCCCATCTCTGTTGACTCGACATGCTGAGAACATTCCTATGGAGACACAGC 3835
 QY ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
 Db 3836 CGGGTGGTGTACAGGAAGAGATCGTGAGGGCAGCAAGAGGAGCCCAACATACATGCTTC 3895
 QY IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
 Db 3896 ATCGAGTCACTGCTTAATAATATACATTAAGTGGAGACAAAGAACTAGCTCTCT 3955
 QY GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
 Db 3956 GGTGGCCAGAAACACGCTGCTGATGCTGCTGGCCCTGTTAGACAGCCTCATATTTTG 4015
 QY LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
 Db 4016 CTTTGGATGAAGCCAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 4075
 QY LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
 Db 4076 CTGGACAAAGCCAGAGAGCCGCTGCTGATGATGATGATGATGATGATGATGATGATG 4135
 QY GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
 Db 4136 CAGAATGCAGACTTAATAGTGGTGTTCAGATGGCAGAGTCAAGAGGATGGCAGCAT 4195
 QY GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279

Db 4196 CACGAGCTGCTGCCACAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 4255
 QY 1280 LysArg 1281
 Db 4256 AAGCGC 4261
 RESULT 14
 AAT13394
 ID AAT13394 standard; DNA; 6505 BP.
 XX
 AC AAT13394;
 XX
 DT 24-JUN-1996 (first entry)
 XX
 DE Hybrid vector pSP-MDR.
 XX
 KW Hybrid; vector; gene transfer; gene therapy; haematopoietic stem cell;
 KW retroviral; murine embryonic stem cell virus; MESV;
 KW Moloney murine sarcoma virus; (MoMuSV);
 KW Friend murine leukaemia virus; F-MuLV; ds.
 XX
 OS Synthetic.
 XX
 PN DE19503952-A1.
 XX
 PD 14-MAR-1996.
 XX
 PF 07-FEB-1995; 95DE-1003952.
 XX
 PR 08-SEP-1994; 94DE-1431973.
 XX
 PA (BOEF) BOEHRINGER MANNHEIM GMBH.
 XX
 PI Baum C, Ostertag W, Stocking-harbers C, Stockingharbers C;
 XX
 DR WPI; 1996-152306/16.
 XX
 PT Hybrid retroviral vectors - for gene transfer into haematopoietic
 PT stem cells
 XX
 PS Disclosure; Page 25-29; 42pp; German.
 XX
 CC New hybrid vectors comprise (1) a leader region including the U5
 CC region and trna primer binding site of murine embryonic stem cell
 CC virus (MESV) or Moloney murine sarcoma virus (MoMuSV), and (2) a 3'-
 CC LTR including the U3 and R regions of a Friend murine leukaemia
 CC virus (F-MuLV). The vectors are useful for ex-vivo or in-vivo gene
 CC therapy. High levels of gene transfer can be achieved in
 CC haematopoietic stem cells and their myeloid (non-lymphatic) progeny.
 CC pSP1, pSF2, pSF3 and pMW1 (sequences given in AAT13390-13393) are
 CC examples of such vectors.
 CC Vector pSP-MDR (sequence given in AAT13394) is based on the
 CC MESV vector R224.
 XX
 SQ Sequence 6505 BP; 1719 A; 1506 C; 1606 G; 1674 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 6505
 Score: 5857.00 Matches: 1163
 Percent Similarity: 95.01% Conservative: 55
 Best Local Similarity: 90.72% Mismatches: 60
 Query Match: 90.48% Indels: 4
 DB: 17 Gaps: 4
 US-09-672-725C-25 (1-1281) x AAT13394 (1-6505)
 QY 1 MetaspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
 Db 1817 ATGGATCTTGAAGGGCCGCAATGGAGGACCAAGAGAGAACTTTTAACTGAAC 1876
 QY 20 LysLysSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 39
 Db 1877 AATAAAAGT---GAAAGAGATAAGAGGAAAGAAAGCACTGTCTGATTTTCAATG 1933

QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
DB 1934 TTTGCGTATTCAAATGGCTTGACAAAGTTGTATATGGTGGGAACCTTGGCTGCCATC 1993
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
DB 1994 ATCCATGGGCGTGGACTTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 2053
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
DB 2054 GCAAAATGCAGGA---AATTTAGAAGATCTGATGTCACAAATCATCAATAGAGTGTATATC 2110
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
DB 2111 AATGATACAGGGTTCTTCATGAAT---CTGGAGGAAGACATGACCAGGTATGCCATTTAT 2167
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaIleTyrIleGlnValSerPheTrpCys 139
DB 2168 TACAGTGAATGGTCTGGGTGCTGGTGTGCTTACATTCAGGTTCATTTGGTGC 2227
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
DB 2228 CTGGCAGCTGGAGACAAATACACAAATAGAAAACAGTTTTTCATGCTATAATGGCA 2287
QY 160 GlnIleIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
DB 2288 CAGGAGATAGGCTGTTGATGTGCACGATGTTGGGGAGCTTAACACCGGACTTACAGAT 2347
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
DB 2348 GATGCTCTAAGATTAAATGAAGTTATTGGTGACAAATTTGGAATGTTCTTTCAGTCAATG 2407
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrVal 219
DB 2408 GCACCAATTTTCTACTGGGTTATAGTAGGATTTACAGTGGTGGAGCTTAACCTTTGTG 2467
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
DB 2468 ATTTTGGCCATCAGTCTGTTCTTGGACTGTGAGCTGCTGCTGGGCAAGATCTATCT 2527
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
DB 2528 TCATTTACTGATAAAGAACTTTAGCGTATGCCAAAGCTGGAGCAGTAGCTGAAGAGGTC 2587
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
DB 2588 TTGGCAGCAATTAAGAACTGTGATGTTGGAGGACAAAGAAAGAACTTGAAGGTAC 2647
QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
DB 2648 AACAAAAATTTAGAAGAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCAATATT 2707
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
DB 2708 TCTATAGTCTGCTCTTCTGCTGATCTATGCTATGCTATGCTGCTGCTGCTGGATGGG 2767
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
DB 2768 ACCACCTTGTCTCTCAGGGGAATATTCTATTGGACAGTACTCACTGATCTTTCT 2827
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
DB 2828 GTATTAAATGGGCTTTTACTGTTGGACAGGCATCTCCAAGCATTGAAGCAATTTGCAAT 2887
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
DB 2888 GCAGAGGACAGCTTATGAATCTTCAAGATAATTGATAAAGCAAGTATTGACAGC 2947
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
DB 2948 TATTCGAAGAGTGGGCACAAACACAGATAATTATTAAGGGAATTTGGAATTCAGAAATGTT 3007

QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
DB 3008 CACTTCAGTTACCATCTCGAAAGAAAGTTAAGATCTTGAAGGCGCTGAACCTGAAGGTG 3067
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
DB 3068 CAGAGTGGGAGAGCGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGACACACAGTC 3127
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
DB 3128 CAGCTGATGACAGGCTCTATGACCCACAGAGGGGATGCTAGTGTGTGACAGGAT 3187
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
DB 3188 ATTAGGACCAATAATGTAAGTTTCTACGGGAAATCATTTGCTGTGGTGAAGTCAAGAACCT 3247
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
DB 3248 GTATTGTTGGCCACCATGATAGCTGAAACATTCGCTATGCGCTGAAATGTCAACCATG 3307
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
DB 3308 GATGAGATTGAGAAAGCTGTCAAGGAAGCCAAATGCCCTATGACTTTATCATGAAACTGCCCT 3367
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
DB 3368 CATAAATTTGACACCCCTGGTTGGAGAGAGGGGCCAGTTGAGTGGTGGGCAAGACAG 3427
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
DB 3428 AGGATGCGCATGTCAGTGGCTGTTGCGAACCCCAAGATCCTCTCTGCTGGATGAGGCC 3487
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
DB 3488 ACCTCAGCTTGGACACAGAAAGCGAAGCAGTGGTTCAGGTGGCTCTGGATAAGGCCAGA 3547
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
DB 3548 AAAGTCGGACCAACCATTTGATAGCTATCGTTTGTCTACAGTTCGTAATGCTGACGTC 3607
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
DB 3608 ATCGCTGTTTCGATGATGGAGTCATTGTGGAGAAAGAAATCATGATGAACATCATGAAA 3667
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
DB 3668 GAGAAAGCATTTACTTCAAACTTGTCAAAATGCAGACAGCAGCAAGAAATGAAGTTGAATTA 3727
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
DB 3728 GAAATGCGAGCTGATGAATCCAAAAGTGAATTCATGCTTGGAAATGCTCTCAAAATGAT 3787
QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
DB 3788 TCAGATCCAGTCTAATAAGAAAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAAGCC 3847
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
DB 3848 CAAGACAGAAAGCTTAGTACCAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 3907
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
DB 3908 TGGAGGATTAAGACATAAATTAACGAATGGCCCTATTTGTTGTTGGTGTATTGTTGT 3967
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
DB 3968 GCCATTATAAATGGAGGCTGCAACACAGCATTTGCAATAAATATTTTCAAGATATATAGG 4027
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
DB 4028 GTTTTACAAGAAATGATGATCCTTGAAACAAACACAGAAATAGTAATCTGTTTCACTA 4087
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:39:50 : Search time 58.556 Seconds
(without alignments)
5373.600 Million cell updates/sec

Title: US-09-672-725C-25

Performance: 6473

Sequence: 1 MDPGGRKGSARFNWKGK.....LLAQGIYFMSVQAGAKR 1281

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DBV=xlh
-Q/cgn2_1/USPto_spool/US09672725/runat_04112002_124339_17887/app_query.fasta_1.6435
-DB=Issued_Patents_NA -QFmap=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725 -CGN1_1_134 -runat_04112002_124339_17887 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:**
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:**
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:**
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:**
- 5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:**
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5857	90.5	4646	1	US-08-181-471-2
2	5857	90.5	6505	2	US-08-793-610-5
3	5857	90.5	9318	2	US-08-793-610-6
4	5854	90.4	4669	6	5206352-3
5	5830	90.1	4669	2	US-08-752-447-1
6	5815.5	89.8	4264	2	US-08-784-649A-1
7	5815.5	89.8	4264	2	US-08-784-649A-5
8	5797	89.6	4669	2	US-08-583-276-18
9	5291.5	81.7	4233	3	US-09-120-513-1
10	5291.5	81.7	4233	3	US-09-450-105-1
11	3628	56.0	2726	1	US-08-461-823-1
12	2513	38.8	4047	2	US-08-612-734B-1

Sequence 2, Appli
Sequence 5, Appli
Sequence 6, Appli
Patent No. 5206352
Sequence 1, Appli
Sequence 1, Appli
Sequence 5, Appli
Sequence 18, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli

13	2481	38.3	4002	2	US-08-996-545-1	Sequence 1, Appli
14	2481	38.3	4002	2	US-08-996-545-3	Sequence 3, Appli
15	2481	38.3	4002	4	US-09-328-320-1	Sequence 1, Appli
16	2481	38.3	4002	4	US-09-328-320-3	Sequence 3, Appli
17	2446	37.8	4224	1	US-08-612-521-1	Sequence 1, Appli
18	2438.5	37.7	4800	2	US-08-612-734B-3	Sequence 3, Appli
19	2288	35.3	3924	1	US-08-395-246C-1	Sequence 1, Appli
20	2059	31.8	6143	1	US-08-612-521-3	Sequence 3, Appli
21	1877.5	29.0	3924	3	US-08-996-644-3	Sequence 3, Appli
22	1877.5	29.0	3924	3	US-09-352-552-3	Sequence 3, Appli
23	1877.5	29.0	3927	3	US-08-996-644-1	Sequence 1, Appli
24	1877.5	29.0	3927	3	US-09-352-552-1	Sequence 1, Appli
25	1855	28.7	3909	1	US-08-232-537-1	Sequence 1, Appli
26	1088	16.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
27	1088	16.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
28	778.5	12.0	2376	1	US-08-394-880B-1	Sequence 1, Appli
29	766	11.8	4781	2	US-09-001-273-1	Sequence 1, Appli
30	766	11.8	4781	4	US-08-843-459A-1	Sequence 1, Appli
31	766	11.8	4847	3	US-09-061-400-1	Sequence 1, Appli
32	753.5	11.6	5889	1	US-08-463-092B-5	Sequence 5, Appli
33	753.5	11.6	5889	2	US-08-462-109A-5	Sequence 5, Appli
34	753.5	11.6	5889	3	US-08-460-907B-5	Sequence 5, Appli
35	753.5	11.6	5889	3	US-08-463-179A-5	Sequence 5, Appli
36	753.5	11.6	5889	3	US-08-461-384B-5	Sequence 5, Appli
37	728	11.2	6140	4	US-09-439-313-536	Sequence 536, App
38	710	11.0	6082	4	US-09-439-313-535	Sequence 535, App
39	708	10.9	5232	4	US-08-972-927-1	Sequence 1, Appli
40	691	10.7	2061	4	US-09-061-764A-17	Sequence 17, Appli
41	685	10.6	5011	1	US-08-463-092B-3	Sequence 3, Appli
42	685	10.6	5011	2	US-08-462-109A-3	Sequence 3, Appli
43	685	10.6	5011	2	US-08-460-907B-3	Sequence 3, Appli
44	685	10.6	5011	3	US-08-463-179A-3	Sequence 3, Appli
45	685	10.6	5011	3	US-08-461-384B-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-181-471-2
; Sequence 2, Application US/08181471
; Patent No. 5641508
; GENERAL INFORMATION:
; APPLICANT: Li, Lingna
; APPLICANT: Lishko, Valeryi K.
; TITLE OF INVENTION: METHOD FOR DELIVERING BENEFICIAL
; TITLE OF INVENTION: COMPOSITIONS TO HAIR FOLLICLES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas Fitting
; STREET: 12526 High Bluff Drive, Suite 300
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92130
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/181,471
; FILING DATE: 13-JAN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/041,553
; FILING DATE: 02-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: ANT0029P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-792-3680
; TELEFAX: 619-792-8477

for clear view are

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4646 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 425..4267

US-08-181-471-2

Alignment Scores:

Pred. No.:	0	Length:	4646
Score:	5857.00	Matches:	1163
Percent Similarity:	95.01%	Conservative:	55
Best Local Similarity:	90.72%	Mismatches:	60
Query Match:	90.48%	Indels:	4
DB:	1	Gaps:	4

US-09-672-725c-25 (1-1281) x US-08-181-471-2 (1-4646)

QY	1	MetAspProGluGlyGlyArgLysGlySerAla--GluLysAsnPheTrpLysMetGly	19
DB	425	ATGGATCTTGAAGGGACCGCAATGGAGGAGCAAGAAAGAACTTTTAACTGAAC	484
QY	20	LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheLamMet	39
DB	485	AATAAAAGT---GAAAAAGATAAGAGGAAAGAAACCAACTGTCAGTGTATTTCAATG	541
QY	40	PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLalile	59
DB	542	TTTCGTATTCAAATTTGGCTTGCACAGTTGTATATGTTGGTGGGAACCTTTGGCTGCATC	601
QY	60	IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe	79
DB	602	ATCCATGGGCTGGACTCTCTCATGATCGCTGGTGTGGAGAAATGACATATCTTTT	661
QY	80	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr	99
DB	662	GCAATATGCAGGA---AATTAGAAAGATCTCATGTCAAAACATCACTAATAAGAGTGATAFC	718
QY	100	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr	119
DB	719	AATGATACAGGGTCTTCATGAAAT---CTGGAGAGACATGACACAGGTATGCCATTAT	775
QY	120	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys	139
DB	776	TACAGTGGAAATGGTGTGGGGTGTGCTGTGCTGTACATTCAGGTTTCAITTTTGGTGC	835
QY	140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159
DB	836	CTGGACGCTGGAAGACAAATACACAAATTTAGAAAAACAGTTTTTTCATGCTAATAATGCCA	895
QY	160	GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179
DB	896	CAGGAGATAGCGTGTGTATGTGCACAGATGTTGGGGAGCTTAACACCCGACTTACAGAT	955
QY	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle	199
DB	956	GATGCTCTTAAGATTAAATGAAGTTATTGGTGACAAAATTTGGAATGTTCTTTCAGTCAATG	1015
QY	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219
DB	1016	GCAACATTTTTCACCTGGGTTTATAGTAGGATTTACACGTGGTGGAGGCTTAACCCCTTGG	1075
QY	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239
DB	1076	ATTTTGGCCATCAGTCTCTTCTTGGACTGTCAGTGTCTGCTGGGCAAGATATCTATCT	1135
QY	240	SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal	259

DB	1136	TCATTTACTATAAAGAACTCTTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC	1195
QY	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluAlaTyrTyr	279
DB	1196	TTGCAGCAATTAGAAGCTGTGATTTGGAGGACAAAGAAAGAACTTGAAGGTAC	1255
QY	280	AsnLysAsnLeuGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle	299
DB	1256	AACAAAATTTAGAAAGAACTAAAAGAAATGGGATAAAGAAAGCTATTACAGCAATATT	1315
QY	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrTyrGly	319
DB	1316	TCTATAGTGCTGCTTTCCTCTGATCATCTATCTTATGCTCTGGCCTTCTGGTATGGG	1375
QY	320	ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer	339
DB	1376	ACCACCTTGGTCTCTCAGGGGAATATCTATTGGACAGACTACACTGATTTCTTTCT	1435
QY	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359
DB	1436	GTATTAAATGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAT	1495
QY	360	AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379
DB	1496	GCAAGAGAGCAGCTTATAAATCTTCAAGATAATTGATAATAAGCCAAAGTATTGACAGC	1555
QY	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399
DB	1556	TATTGGAAGAGTGGCACAACCAAGATAATTAATTAAGGGAATTTTGGAAATTCAGAAATGTT	1615
QY	400	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419
DB	1616	CACITTCAGTTACCCATCTCGAAAAAGAGTAAAGATCTTGAAGGGCCCTGAACCTGAAGGTG	1675
QY	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal	439
DB	1676	CAGAGTGGCAGACGGTGGCCCTGTTGGAAACAGTGGCTGTGGGAAGAGCACCAAGTC	1735
QY	440	GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459
DB	1736	CAGCTGATGACAGAGCTCTATGACCCACAGAGGAGTGGTCACTGTTGATGGACAGGAT	1795
QY	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro	479
DB	1796	ATTAGCACATAAATGTAAGGTTTCTACGGGAAATCATTTGTTGGTGGTCAAGAACCT	1855
QY	480	ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499
DB	1856	GTATTGTTTGGCCACCATAGCTGAAACAAATTCGTTATGGCCCTGAAATGTCCACATG	1915
QY	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro	519
DB	1916	GATGAGATTGAGAAAGCTGTCAAGGAAGCCCAATGCCTATGACTTTTATCATGAAACTGCCT	1975
QY	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539
DB	1976	CATAAATTTGACACCTCGTGGTGGAGAGAGAGGGGCCCACTGAGTGGTGGGCAAGACAG	2035
QY	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559
DB	2036	AGGATGGCCATTGCACGTGCCCTGGTTCGCAACCCCAAGATCCCTCCTGCTGGATGAGGCC	2095
QY	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579
DB	2096	AGCTCAGCCTTGGACACAGAAAGCAAGCAGTGGTTCAGTGGCTCTGGATAAGGCCAGA	2155
QY	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599
DB	2156	AAAGTCGGACCAACCATTTGTGATAGCTCATCTGTTGCTACAGTTCTGTAATGCTGACGTC	2215
QY	600	IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619

Db 2216 ATCGCTGGTTTGGATGATGGAGTCAATGTGGAGAAAGGAAATCATGTAACATCATGAAA 2275
Qy 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 2276 GAGAAGGCATTTACTTCAACTTGTCAATGCAGACAGCAGGAAATGAAGTTGAATTA 2335
Qy 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2336 GAAAATGCAGCTGATGAATCCAAAGTGAATTTGATGCTTGGAAATGCTCTCAATGAT 2395
Qy 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleIleAlaProGlnGly 679
Db 2396 TCAAGATCCAGTCTAATAAGAAAGATCAACTCGTAGAGTGTCCGTGGATCAACAAGCC 2455
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGlnAsnValProProValSerPhe 699
Db 2456 CAAGACAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 2515
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2516 TCGAGGATTATGAAGCTAAATTAACGATGGCTTATTTTGTGTGGTATTGTTGT 2575
Qy 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2576 GCCATTATAAATGGAGGCTCGCAACAGCAATTCGAATAATATTTTCAAAAGATTATAGG 2635
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 GTTTTTACAGAANTGATGATCCTGGAACAAACAGCAGAAATAGTAATCTGTTTTCCTA 2695
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2696 TTGTTTCTAGCCCTTGGAAATATTTCTTTATTAATTTTCTTCAGGGTTTCACATTT 2755
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 GGCAAAAGCTGGAGAGATCCTCACCAAGCGCTCCGATACATGTTTTCGATCCATGCTC 2815
Qy 800 ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2816 AGACAGATGTGAGTTGGTTGATGACCCCTAAACACCACTGGAGCATTTGACTACCAGG 2875
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleIleGlySerArgLeuAlaValIleThr 839
Db 2876 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCAGGCTTGCTGTAAATTACC 2935
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTyrGlnLeu 859
Db 2936 CAGAAATAGCAATCTTGGGACAGGAATAATTATATCCTTCATCTATGTTGGCAACTA 2995
Qy 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2996 ACATGTTACTCTTAGCAATGTACCCATGTTGCAATAGCAGAGTTGTTGAATGAAA 3055
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 3056 ATGTTGTCTGGCAAGCACTGAAGATAAGAAAGAACTAGAAAGTGTGGGAGATCGCT 3115
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 3116 ACTGAGCAATAGAAAATCTCCCAACCGTTTCTTCTTGTGACTCAGGACAGAGATTGAA 3175
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGTATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAGACACATC 3235
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTTGGAAATACATTTTCTTCACCCAGGCAATGATGATATTTTCTTATCTGATGATGTTTC 3295
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 3296 CGGTTTGGAGCCCTACTTGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 3355

Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3356 TTTTCAGCTGTGCTCTTGTGTGCATGGCCGTGGGCAAGTCAGTTCAATTTGCTCCTGAC 3415
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAAGCCAAATATCAGCAGCCACATCATGATCATTTGAAAACCCCTTTG 3475
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTGACAGCTACAGCAGCGAAGCCTAATGCCGAACACATTTGGAAGAAATGTCACATT 3535
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGATTTCAACTATCCACCCGACCGGACATCCAGTGTCTCAGGACTGAGC 3595
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3596 CTGGAGGTGAAGAGGCCAGAGCGCTGGCTCTGGTGGGAGCAGTGGCTGTGGGAAGAC 3655
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3656 ACAGTGTCTCAGCTCTCGAGCGGTTCTACGACCCCTTGGCAGGGAAGTGTCTGTGAT 3715
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 GGCAAAAGAAATAAAGGACTGAATGTTCAGTGGCTCCGAGCACACCTGGGCTCGTGTCC 3775
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 CAGAGCCCATCTCTGTTGACTGCACATTTGCTGAGAACATTTGCCCTATGGACACAGC 3835
Qy 1140 ArgValValSerHisGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
Db 3836 CGSGTGTGTACAGGAAGAGATCGTGGAGGAGCAAGAGAGCCCAACATACATGCCCTC 3895
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCGATGCTACTGCCTTAATAATATAGCACTAAGTAGGAGACAAAGAACTCAGCTCT 3955
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GGTGGCCAGAAACAGCAATGCCATAGCTCGTGGCTTGTAGACGCTCATATTG 4015
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAGCC 4075
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGGCAAAAGCCAGAGAGGCCACCTGCAATTTGATGCTCACCCTGCTCCACCATC 4135
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGATGCAGACTTAATAGTGTGTTTCAAGATGCAGAGTCAAGAGCATGGCAGCAT 4195
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGTGGCAGAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 4255
Qy 1280 LysArg 1281
Db 4256 AAGCGC 4261

RESULT 2

US-08-793-610-5
; Sequence 5, Application US/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARBERS, Carol
; APPLICANT: OSTERTAG, Wolfram

;; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
;; TITLE OF INVENTION: FOR GENE TRANSFER
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
;; STREET: 655 Fifteenth Street N.W. Suite 330
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20005-5701
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/793,610
;; FILING DATE: 07-MAR-1997
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: DE P 44 31 973.8
;; FILING DATE: 08-SEP-1994
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: DE 195 03 952.1
;; FILING DATE: 07-FEB-1995
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: PCT/EP95/03175
;; FILING DATE: 10-AUG-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Berman, Richard J.
;; REGISTRATION NUMBER: 39,105
;; REFERENCE/DOCKET NUMBER: P1614-7007
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)638-5000
;; TELEFAX: (202)638-4810
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6505 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: circular
;; MOLECULE TYPE: DNA
;; US-08-793-610-5

Alignment Scores:
Pred. No.: 0 Length: 6505
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 2 Gaps: 4

US-09-672-725C-25 (1-1281) x US-08-793-610-5 (1-6505)

Qy 1 MetaspProGluGlyArgLysGlySerAla---GluLysAsnPhetrPlysMetGly 19
Db 1817 ATGGATCTTGAAGGGCCGCAATGGAGGAGCAAGAAGAACTTTTAACTGAAC 1876
Qy 20 LysLysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 1877 AATAAAAGT---GAAAAGATAGAAGAAAGAAACCAACTGCTAGTGTATTTCAATG 1933
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAla 59
Db 1934 TTTTCGCTATTCAAAATGGCTTGACAAGTGTATATGGTGGGAACTTTGGCTGCCATC 1993
Qy 60 IleHsGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 1994 ATCCATGGGCTGGACTTCCTCATGATGCTGGTGGAGAAATGACAGATATCTTT 2053
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 2054 .GCAATGCAGGA---AATTAGAAGATCTGATGTCAACATCATCAATAAGAGTGATATC 2110

Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 2111 AATGATACAGGGTCTTTCATGAAT---CTGAGAGACACATGACCAGGTATGCTATTAT 2167
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 2168 TACAGTGAATTTGGTGGTGGTCTGCTGCTTACATTGAGGTTTCATTTGGTGC 2227
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 2228 CTGCAGCTGGAAGACAAATACACAAAATAGAAAACAGTTTTCATGCTATAATGCGA 2287
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 2288 CAGGAGATAGCTGTTTGTATGTCACGATGTTGGGAGCTTAACACCCGACTTACAGAT 2347
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 2348 GATGCTCTAAGATTAATGAAGTTATTGGTGACAAAATTTGGAATGTTCTTTCAGTCAATG 2407
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 2408 GCAACATTTTTCAGTGGTATTATAGTAGATTACACGTGGTTGGAAGCTAACCCTTGTG 2467
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 2468 ATTTTGGCCATCAGTCCTGTTTGGACTGTCAGCTGCTGCTGGGCAAGATACTATCT 2527
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
Db 2528 TCATTTACTGATAAAGAACTCTTAGCGGTATGCAAAAGCTGGAGCAAAAGAACTTGAAGGTAC 2587
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 2588 TTGGCAGCAATTAGAACTGCTGATTGCAATTTGGAGGCAAAAGAAAGAACTTGAAGGTAC 2647
Qy 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 2648 AACAAAAATTTAGAAGAACTTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCAAATT 2707
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 2708 TCTATAGTGTGCTTTCCTGCTGATCATCTATGCTTATGCTCTGGCTTCTGATATGGG 2767
Qy 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 2768 ACCACCTTTGGTCTCTCAGGGCAATATTCTATTGGACAAGTACTCCTGTTATCTTTCT 2827
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 2828 GTATTTAATTTGGGGCTTTTAGTGTGGACAGGCATCTCCAACGANTTGAAGCATTTGCAAT 2887
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 2888 GCACAGAGAGCAGCTTATGAATCTTCAAGATAATTGATAATAAGCCAAAGTATTGACAGC 2947
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 2948 TATTCGAAGAGTGGGCACAAACACAGATAATATTAAAGGAAATTTGGAATTCAGAAATGTT 3007
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 3008 CACTTCAGTTACCCATCTCGAAAAGAAAGTATAGATCTTGAAGGGCCTGAACCTGAAGGTG 3067
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 3068 CAGAGTGGGCACAGCGTGGCCCTGTTGGAACAGTGGCTGTGGAGAGAGACACACAGTC 3127
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 3128 CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGGATGGTCAGTGTGTATGGACAGGAT 3187
Qy 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479

Db 3188 ATTAGGACCAATAAATGAAGTTTCTACGGGAAATCATTTGGTGTGGAGTCAGGAACCT 3247
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 3248 GTATTGTTGGCCACCAGATAGCTGAAACAATTCGCTATGGCCGTGAAAAATGTCCACATG 3307
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 3308 GATCAGATTGAGAAGCTGTCAAGAAAGCCATGCGCTTATGACTTTATCATGAACCTGCCT 3367
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 3368 CATAAATTGACACCTTGGTGGAGAGAGAGGGCCAGTTGAGTGGTGGCAGAGCAG 3427
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAla 559
Db 3428 AGGATCGCCATTGACGTGCGCTGGTTCCGAACCCCAAGATCCTCCTGCTGGATGAGGCC 3487
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 3488 ACCTCAGCCTTGGACACAGAAAGCAGAGTGGTTCAGGTGGCTCTGGATAAGGCCAGA 3547
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 3548 AAAGTCGGACCACCATTTGTATAGCTCATCGTTTGTCTACAGTTCTGTAATGCTGACGTC 3607
QY 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisaspGluLeuMetLys 619
Db 3608 ATCGCTGGTTCGATGATGAGTGCATTTGTGGAGAAAGAAATCATGATGAACCTCATGAA 3667
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 3668 GAGAAAGGCATTTACTTCAACTTGTCCACAAATGCACACAGCAGAAATGAAGTTGAATTA 3727
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 3728 GAAATGCAGCTGATGAATCCAAAAGTGAATTTGATGCGCTTGGAAATGCTTCAAAATGAT 3787
QY 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 3788 TCAAGATCCAGTCTAATAAGAAAGATCAACTCGTAGGAGTGTCCGTGGATCATCAAGCC 3847
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
Db 3848 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 3907
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 3908 TGGAGGATTATGAAGCTAAATTTAACTGAATGGCCCTATTATTGTTGTTGTTGTTATTTGT 3967
QY 720 AlaIleIleAsnGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 3968 GCCATTATAATGGAGCCCTGCACACAGCATTTGGCAATAATATTTTCAAGATTTATAGG 4027
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 4028 GTTTTACAGAATTGATGATCCTGAAACAAACACAGACAATAGTANCTGTGTTTCACTA 4087
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 4088 TTGTTTCTAGCCCTTGGAAATTTATTTTATTATATTTTCTTCTTCAGGGTTTCACATTT 4147
QY 780 GlyLysAlaGlyLutIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 4148 GGCAAAGCTGGAGAGATCTCCACCAAGCGCTCCGATACATGGTTTCCGATCCATGCTC 4207
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 4208 AGACAGGATGTGAGTTGGTTGATGACCTTAAACACACCACCTGGAGCATTTGACTACCAGG 4267
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839

Db 4268 CTGCGCCATCATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTCTGTAATTACC 4327
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 4328 CAGATATAGCAAAATCTTGGCAGAGATAATATATATCTTCTATCTATGTTGGCAACTA 4387
QY 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 4388 ACACTGTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAA 4447
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 4448 ATGTTGCTGGCAGACACTGAAAGATAGAAGAACTAGAAGGTGCTGGGAGATCGCT 4507
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 4508 ACTGAAGCAATAGAAAACCTTCGAAACCGTTGTTCTTCTGACTCAGGAGCAGAAAGTTGAA 4567
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 4568 CATATGTATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAACACACATC 4627
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 4628 TTTGGAATTTACATTTTCTTCCACCCAGGCAATGATGATATTTTCTATGCTGGATGTTT 4687
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 4688 CGGTTTGGAGCCTACTTGGTGCACATAAACTCATGAGCTTTGAGGATGTTCTGTTAGTA 4747
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 4748 TTTTCAGCTGTGTCTTGGTGCCATGGCCGGGGGAGTCAGTTCATTTGCTCCTGCAC 4807
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 4808 TATGCCAAAGGCCAAATATATCAGCAGCCACATCATCATGATCATTTGAAAAACCCCTTG 4867
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 4868 ATTGACAGCTACAGCAGCGAAGGCCCTAATGCCGACACATTTGGAAGGAAATGTACATTT 4927
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 4928 GGTGAAGTTGTATTCACTATCCACCCAGCGACATCCAGAGCTTCCAGGACTGAGC 4987
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 4988 CTGGAGGTGAAGAGGGCCAGACGCTGGCTCTGGTGGGCAGCAGTGGCTGTGGGAAGAGC 5047
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 5048 ACAGTGGTCCAGCTCCTCGAGCGGTTCTACGACCCCTTTGGCAGGGAAGAGTCTGTTGAT 5107
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 5108 GCAGGAGGAGGAGGAGTCTTACGACGCTTACGACCCCTTTGGCAGGGAAGAGTCTGTTGAT 5167
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 5168 CAGGAGCCCATCTGTTTGAAGTGCAGCATTTGCTGAGAACAATTTGCTATGGAGACAACAGC 5227
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPhe 1159
Db 5228 CGGTGGTGTCCAGGAGAGAGATCTGTGGGCGAGCAAGGAGGCAACATACATGCTCCTTC 5287
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 5288 ATCAGTCACTGCTTAATAATATAGCACTAAAGTAGAGCAAGGAACTCAGCTCTCT 5347
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 5348 GGTGGCCAGAAACACGATTTGCCATAGCTGCTGCCCTTGTGTAGACAGCCTCATATTTTG 5407

QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 5408 CTTTGGATGAAGCCAGCTGAGCTGGATACAGAAAGTGAAGAGTTGTCCAGAACCC 5467
QY 1220 LeuAspLysAlaArgGluGluArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 5468 CTGGACAAAGCCAGAGAGCCGACCTGATTTGATTCACCGCTGTCCACCATC 5527
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 5528 CAGAATGCAGACTTAATAGTGTGTTTCAGATGCGAGAGTCAAGGAGCATGGCAGCAT 5587
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 5588 CAGCAGCTGCTGGCAGAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 5647
QY 1280 LysArg 1281
Db 5648 AAGCGC 5653

RESULT 3
US-08-793-610-6
; Sequence 6, Application US/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARBERS, Carol
; APPLICANT: OSTERTAG, Wolfram
; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
; TITLE OF INVENTION: FOR GENE TRANSFER
; NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:
ADDRESSEE: Nikolaïdo, Marmelstein, Murray & Oram LLP
STREET: 655 Fifteenth Street N.W. Suite 330
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-5701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,610
FILING DATE: 07-MAR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 44 31 973.8
FILING DATE: 08-SEP-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 195 03 952.1
FILING DATE: 07-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/03175
FILING DATE: 10-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Beiman, Richard J.
REGISTRATION NUMBER: 39,105
REFERENCE/DOCKET NUMBER: P1614-7007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)638-5000
TELEFAX: (202)638-4810
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 9318 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA
US-08-793-610-6

Alignment scores:

Pred. No.: 0 Length: 9318
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservatives: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 2 Gaps: 4
US-09-672-725c-25 (1-1281) x US-08-793-610-6 (1-9318)
QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1776 ATGGATCTTGAAGGGGACCCTGAGGAGCAAGAGAGAACTTTTAACTGAAC 1835
QY 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 1836 AATAAAGT---GAAAGAGATAGAGAGAAACCAACTGTCAGTGTATTTCAATG 1892
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
Db 1893 TTTGCTATTCAAATTTGGCTTGACAGTTGATATGTTGGTGGAACTTTGGCTGCCATC 1952
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 1953 ATCCATGGGCTGGACTTCTCTCATGATGCTGTGTGGAGAAATGACAGATATCTTT 2012
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 2013 GCAATGCAGGA---AATTTAGAGATCTGATGTCAACATCACTAATAGAGTGATATC 2069
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 2070 AATGATACAGGGTTCTTCATGAAT---CTGGAGGAGACATGACCCAGGTATGCCATATAT 2126
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 2127 TACAGTGAATTTGGTGTGGGCTGCTGGTCTCTTACATTACAGTTTCATTTGGTGC 2186
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 2187 CTGGCAGCTGCAAGACAAATACAAAAATAGAAAAACAGTTTTTTTCATGTATATAATCGA 2246
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 2247 CAGGAGATAGCTGTTGATGTGCAGCATTTGGGAGCTTTAACACCCGACTTACAGAT 2306
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db 2307 GATGCTCTAAGATTATGAAGTTATTTGGTGACAAAATTTGGAATGTTCTTTCACTCAATG 2366
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 2367 GCAACATTTTCACTGGGTTTATAGTAGGATTTACACGTGGTTGGAAAGTAAACCTTTGTG 2426
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 2427 ATTTTGGCCATCAGTCCTGTTCTTGGACTGTCAGCTGCTGCTGGGCAAGAGTACTATCT 2486
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
Db 2487 TCATTTACTATAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTGC 2546
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyr 279
Db 2547 TTGGCAGCAATTAGAACTGTGATTTCATTTGGAGGACAAAAGAACTTTGAAGGTAC 2606
QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 2607 AACAAAAATTTAGAAAGCTTAAAGAAATTTGGATAAAGAAAGCTTATACAGCCAATATT 2666
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 2667 TCTATAGTGTGCTTTCTCTGCTGATCATGATCTTATGCTGCTGGCCTTCTGTGTGGG 2726

Db	4887	GGTGAAGTTGTATTCAACTATCCACCAGCGGACATCCCGAGTGCCTTCAGGGAGCTGAGC	494
Qy	1060	LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer	1079
Db	4947	CTGGAGGTGAAGAGGGCCAGACGCTGGCTCTGTGGCAGCAGCTGGCTGTGGGAAGAGC	5006
Qy	1080	ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp	1099
Db	5007	ACAGTGTGTCCAGCTCTCTGGAGCGGTTCTACGACCCCTTGGCAGGGAAGTCTGCTTGAT	5066
Qy	1100	GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer	1119
Db	5067	GGCAAGAAATAAAGGAGCTGAATGTTCACTAGTGGCTCCGAGCACACCTGGGCATCGTCTCC	5126
Qy	1120	GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer	1139
Db	5127	CAGGAGCCCATCTGTTTGACTGACATGCTGAGAACATGCTATGGAGACACAGC	5186
Qy	1140	ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe	1159
Db	5187	CGGTGTGTGCACAGGAAGAGATCGTGGAGGCGAGCAAGAGGAGGCCAACATACATGCCTTC	5246
Qy	1160	IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer	1179
Db	5247	ATCGAGTCACTGCCTAATAATATAGCACTAAAGTAGGAGACAAGAACTCAGCTCTCT	5306
Qy	1180	GlyClyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu	1199
Db	5307	GGTGGCCAGAAACACGCAATGCCATAGCTGCGCCTTGTAGACAGCCTCATATTTTG	5366
Qy	1200	LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla	1219
Db	5367	CTTTTGGATCAAGCCAGCTCAGCTCGGATACAGAAAGTGAAGAGTTGTCCAAGAGCC	5426
Qy	1220	LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle	1239
Db	5427	CTGGACAAGCCAGAGAAGCGCGACCTGCATTGTGATTGCTCACCGCGCTGTCCACCATC	5486
Qy	1240	GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis	1259
Db	5487	CAGAATGCAGACTTAATAGTGGTGTTCAGAATGCAGATCAAGGAGCATGGCAGCAT	5546
Qy	1260	GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla	1279
Db	5547	CAGCAGCTGTCGCACAGAAAGCATCATTTTTCATAGGTCAGTCTCAGGCTGGGAACA	5606
Qy	1280	LysArg 1281	
Db	5607	AAGCGC 5612	
RESULT 4			
5206352-3			
; Patent No. 5206352			
; APPLICANT: Roninson, Igor B.; Pastan Ira H.; Gottesman,			
; Michael M.			
; TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA			
; SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS			
; NUMBER OF SEQUENCES: 4			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/07/622,836			
; FILING DATE: 24-SEP-1990			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: 892,575			
; FILING DATE: 01-AUG-1986			
; APPLICATION NUMBER: 845,610			
; FILING DATE: 28-MAR-1986			
; SEQ ID NO: 3:			
; LENGTH: 4669			
5206352-3			
Alignment Scores:			
Pred. No. 0			
Score: 5854.00			
Length: 4669			
Matches: 1162			

Percent Similarity:	95.01%	Conservative:	56
Best Local Similarity:	90.64%	Mismatches:	60
Query Match:	90.44%	Indels:	4
DB:	6	Gaps:	4
US-09-672-725C-25 (1-1281) x 5206352-3 (1-4669)			
Qy	1	MetAspProGluGlyGlyArgLysGlySerAla--GluLysAsnPheTrpLysMetGly	19
Db	425	ATGGATCTTGAAGGGGCCCAATGGAGGACCAAGAGAACACTTTTAACTGAAC	484
Qy	20	LysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheLamMet	39
Db	485	AATAAAAGT--GAATAAGATAGAAGGAAAGAAACCAACTGCTAGTGTATTTCAATG	541
Qy	40	PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLalle	59
Db	542	TTTCGCATTCAAATGGCTTGACAAGTGTATATGGTGGTGGAACTTTTGCTGCCATC	601
Qy	60	IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspPhe	79
Db	602	ATCCATGGGGCTGGACTTCCTCATGATGCTGGTGTGGGAAATGACAGATATCTTT	661
Qy	80	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleLeuAsnGluSerIleThr	99
Db	662	GCAAAATCGAGGA--AATTTAGAAGATCTGATCTCAACATCACCATAAGAAGTATATC	718
Qy	100	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluLysMetThrThrTyrAlaTyrTyr	119
Db	719	AATGATACAGGGTCTTCATGAAT--CTGGAGGAAGACATGACCCAGGTATGCCATTAT	775
Qy	120	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys	139
Db	776	TACAGTGGAAATGGTCTGGGGTCTGGTTCGTCCTACATTCAGGTTTCATTTGGTGC	835
Qy	140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159
Db	836	CTGSCAGCTGGAAGACAAATACACAAAATTAGAAAAACAGTTTTCATGCTATAATGCGA	895
Qy	160	GlnGluIleGlyTrpPheAspValHisAspValGlyLeuAsnThrArgLeuThrAsp	179
Db	896	CAGAGATAGGCTGGTTGATGTCAGATGTTGGGAGCTTAACACCCGACTTACAGAT	955
Qy	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle	199
Db	956	GATGCTCTAAGATTAACTCAAGTTATTGGTCACAAAATTTGGAAATGTTCTTCAGTCAATG	1015
Qy	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219
Db	1016	GCAACATTTTTCAGTGGGTTTATAGTAGGATTTTACACGTGGTTGGAAGCTAACCCCTGTG	1075
Qy	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239
Db	1076	ATTTTGGCCATCAGTCCGTCTGTGGACTGTCAGCTGCTGTGGGCAAGATATCTACT	1135
Qy	240	SerPheThrAspLysGlnLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal	259
Db	1136	TCAATTACTGATAAGAACTCTTTAGCGTATGCMAAACCTGGAGCAGTACCTGAAAGGTC	1195
Qy	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr	279
Db	1196	TTGSCAGCAATTAGACTGTGATTGCAATTTGGAGGACAAAGAAAGAACTTGAAAGGTAC	1255
Qy	280	AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299
Db	1256	AACAAAAATTTAGAAAGACTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCCAATATT	1315
Qy	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319
Db	1316	TCATAGGTGCTGCTTCCCTGCTGATCTATGCAATCTTATGCTCTGCCCTTCGGTATGGG	1375
Qy	320	ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer	339

Alignment Scores:	
Pred. No.:	0
Score:	5854.00
Length:	4669
Matches:	1162

Db 1376 ACCACCTTGGTCCTCTCAGGGGAATATCTATTGGACAAGTACTCACCTGTATTCTTTCT 1435
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1436 GTATTAAATGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCCAAT 1495
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1496 GCAAGAGGAGCAGCTTATGAATCTTCAAGATAATTGATAATAGCCCAAGTATTGACAGC 1555
QY 380 TyrSerLysSerGlyHisLysProAspAsnLysGlyAsnLeuGluPheLysAsnVal 399
Db 1556 TATTCGAAGAGTGGGCACAAACAGATAATATTAAGGGAATTTGGAATTCAGAAATGTT 1615
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1616 CACTTCAGTTACCATCTCGAAAGAGATTGAATCTTGAAGGGCCTGAACCTGAAGGTG 1675
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1676 CAGAGTGGCAGACGGTGGCCCTGGTTGAAACAGTGGCTGTGGGAAGAGCACAAAGTC 1735
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1736 CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGGATGGTCAGTGTGATGGACAGAT 1795
QY 460 IleArgThrIleAsnValArgHisLeuArgGlnIleThrGlyValValSerGlnGluPro 479
Db 1796 ATTAGGACCAATAATGTAAAGTTTCTACGGGAATCATTTGGTGGTGGTGCAGGAACCT 1855
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1856 GTATTGTTGGCCACCACGATAGCTGAAACATTTCGCTATGSCCGTGAAATGTCCACATG 1915
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1916 GATGAGATTGAGAAGCTGTCAAGGAAGCCCAATGCCCTATGACTTTATCATGAACCTGCCT 1975
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1976 CATAAATTTGACACCTTGGTGGAGAGAGAGGGGCCAGTTGAGTGGTGGCGAAGCAG 2035
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuLeuAspGluAla 559
Db 2036 AGGATCGCAATGCACGTGCCCTGGTTCGCAACCCCAAGATCCTCCTGCTGGATGAGGCC 2095
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 2096 ACGTCAGCCTTGGACACAGAAAGCGAAGCAGTGGTTCAGGTGGCTCTGGATAAGGCCAGA 2155
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 2156 AAAGSTCGGACCACCATTTGTATAGCTCATCGTTTGTACAGTTCTGTAATGCTGACGTC 2215
QY 500 IleAlaGlyPheAspGlyValIleValIleLysGlyAsnHisAspGluLeuMetLys 619
Db 2216 ATCGCTGGTTTCGATGATGAGTCATCTGTGGAGAAAGAAATCATGATGAACATCAAGAA 2275
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 2276 GAGAAAGGCATTACTTCAAACTTGTACAAATGCAGACAGCAGAAATGAAGTTGAATTA 2335
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2336 GAAATCCAGCTGATGAATCCAAAGTGAATTTGATGGCTTGGAAATGCTTCCAATGAT 2395
QY 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2396 TCAAGATCCAGTCTAATAAGAAAAGATCAACTCGTAGGAGTGTCCGTGGATCCACAGCC 2455
QY 680 GluAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2456 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 2515

QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCys 719
Db 2516 TGGAGGATTATGAAGCTAAATTTAACTGAATGGCTTATTTCTGTGTGTGTATTTGT 2575
QY 720 AlaIleIleAsnGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2576 GCCATTATAATGAGCGCCTGCAACCAAGCATTTGCAATAATATTTTCAAGAGATTATAGG 2635
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 GTTTTACAGAAGATTGATGATCTGAAACAAACAGACAGAAATAGTAACTTGTTCACATA 2695
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2696 TTGTTTCTACCCCTTGGAAATATTTCTTTTATTACATTTTCTTCCAGGGTTTCACATTT 2755
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 GGCAAAGCTGGAGAGATCTTCACCAAGGGCTCCGATACATGTTTTCGATCCATGCTC 2815
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2816 AGACAGGATCTGAGTTGGTTGATGACCTTAAACACACCACCTGGAGCATTTGACTACCAG 2875
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2876 CTGCGCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTCTGTAATTACC 2935
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2936 CAGAATATAGCAATCTTGGCAGAGAAATATATATCTTCATCATGTTGGCAACATA 2995
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2996 ACACCTGTTACTCTTAGCAATTTGACCATCATTTGCAATAGCAGGAGTTCTTGAAATGAA 3055
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 3056 ATGTTGCTGGACAGACACTGAAAGATAAGAAAGAACTAGAAGGTGCTGGGAAGATCGCT 3115
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 3116 ACTGAAGCAATAGAAACTTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAGTTGAA 3175
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGATGCTCAGAGTTTGAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATC 3235
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTGGAATTTACATTTTCTTCCACCGAGCAATGATGATATTTTCTATGCTGGATGTTTC 3295
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 3296 CGGTTTGGAGCCTACTTGTGGCACATAAATCAATGAGCTTTGAGAGTGTCTGTAGTA 3355
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3356 TTTTTCAGCTGTGTCTTTTGGTCCCATGCGGTGGGCAAGTCAGTTTCATTTTCTCCTGCAC 3415
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAAGCCAAATATATCAGAGCCCATCATCATGATCATTTGAAACACCCCTTTG 3475
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTGACAGCTACAGCACGGAAGCCCTAATGCCGAACACATTTGGAAGGAAATGTACATTT 3535
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGTTATTCAACTATATCCACCCGACCGGACATCCAGTCTTTCAGGGACTGAGC 3595

Db 956 |||||GATGTCCTCAAGATTAAATGAAGGAATTGGTCACAAAATTGGAATGTTCTTCAGTCAATG 1015
QY 200 AlathrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 1016 GCAACATTTTTCACCTGGGTTTATAGTAGGATTTACACGTGGTTGGAAGCTAAACCTTTGTG 1075
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 1076 ATTTTGGCCATCAGTCTGTTCTGGACTGTCAGCTGCTGTCGGGCAAGATACTATCT 1135
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 1136 TCATTTACTGATAAAGAACTCTTAGCGTATGCATAAAGCTGGAGCAGTAGCTGAAGAGGTC 1195
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 1196 TTGCACCAATTAGACCTGATTGCAATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 1255
QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 1256 AACAAAAATTAGAAAGACTAAAGAAATTGGGATAAAGAAAGCTATTACAGCCAAATAT 1315
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrGly 319
Db 1316 TCTATAGGTGCTGTTCTCTGCTGATCTATGCATCTTATGCTGTGGCCTTCTGSTATGG 1375
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPheSer 339
Db 1376 ACCACTTGGTCTCTCAGGGGAATATTTCTATTTGGACAGTACTCACTGTATTTCTTTCT 1435
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1436 GTATTAAATGGGCTTTAGTGTGGACAGGCATCTCCAAGCATTTGAGCATTTGGCAAT 1495
QY 360 AlaArgGlyAlaIaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1496 GCAAGAGGACAGCTTATGAAATCTTCAAGATAATTGATAATAAGCAAGTATTGACAGC 1555
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1556 TATTCGAAGAGTGGGCACAAACAGATAATATTAAAGGGAATTTGGAATTCAGAAATGTT 1615
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1616 CACTTCAGTTACCACTCTCGAAGAAGTTAAGATCTTGAGGGCCTGAACCTGAAGAGTG 1675
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1676 CAGAGTGGGCAGACGGTGGCCCTGGTTGGAACAGTGGCTGTGGGAAGAGACACACAGTC 1735
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1736 CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGGATGGTCAGTGTGATGGACAGGAT 1795
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1796 ATTAGGACCATAAATGAAGTTTCTACGGGAATCATTTGGTGTGTGTGAGTCAGGAACT 1855
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1856 GTATTGTTGCCACCACGATAGCTGAAAAACATTGCGTATGGCGTGAAAAATGTCACCATG 1915
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1916 GATGAGATTGAGAAAGCTGTCAGGAAGGCCAATGCCCTATGACTTTATCATGAACCTGCC 1975
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1976 CATAAATTTGACACCTGGTTGGAGAGAGAGGGGCCCACTTGAGTGTGTGGGAGAGCAG 2035
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
|||||

Db 2036 AGGATCGCCATTGTCACGTGCCCTGGTTGCGAACCCCAAGATCCTCTCTGCTGGATGAGGCC 2095
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 2096 ACCTCAGCCTTGGACACAGAAAGCAAGCAGTGGTTTCAGTGGCTCTGATAAGGCCAGA 2155
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 2156 AAAGTGGGACCAACCATTGTGATGACTCATCGTTTGTGTACAGTTGCTGAATGCTACGTC 2215
QY 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 2216 ATCGCTGGTTTCGATGATGAGTCATTGTGGAGAAAGAAATCATGATCACTCATGAA 2275
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 2276 GAGAAAGGCAATTTACTTCAAACTGTGCACATGCACACAGCAGAAATGAAGTTGAATTA 2335
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2336 GAAATGACAGCTGATGAATCCAAAAGTGAAATGTAGTGCCTTGGAAATGCTTCAATGAT 2395
QY 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2396 TCAAGATCCAGTCTAATAAGAAAAAGATCACTCGTAGGAGTGTCCGTGGATCAACAGCC 2455
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2456 CAACACAGAAAGCTAGTACCAGAGGCTCTGGATGAAAGTATACCTCCAGTTTCTCTT 2515
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2516 TGGAGATTATGAAGCTAAATTAATGAATGGCCTTATTTGTTGTTGGTGTATTTGT 2575
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2576 GCCATTATAATGGAGGCTGCAACAGCAGCATTTGCAATAATATTTTCAAAGATTATAGG 2635
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 GTTTTACAGAATTTGATGATCCTGAAACAAACACAGAAATAGTGTGTTTTCACATA 2695
QY 760 LeuPheLeuValIleGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2696 TTGTTTCTAGCCCTTGGAAATTAATTTCTTTATATCATTTTCTTCAGGGTTTCACATTT 2755
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 GGCAAGCTGGAGAGATCCTCACCAAGCGCTCCGATACATGTTTCCGATCCATGCTC 2815
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2816 AGACAGGATGTGAGTTGGTTTCATGCCCTAAAAACACCACTGGAGCATTTGACTACCAG 2875
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2876 CTCGGCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGCCTTGCTGTAAATACC 2935
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2936 CAGAATATAGCAAACTCTGGGACAGCAATAATATATCTTCACTATGTTGGCAACTA 2995
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2996 ACACGTGTACTCTTAGCAATTTGACCCCATCATTCATATAGCAGGAGTGTGTTGAATGAAA 3055
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluAlaGlyLysIleAla 899
Db 3056 ATGTTTGTGTCAGACAGCACTGAAAGATAGAAGAACTAGAAAGGTGCTGGGAAGATCGCT 3115
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 3116 ACTGAAGCAATAGAAAACTTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAGTTTGAA 3175
|||||

QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGTATGCTCAGAGTTTGAGGTACCATACAGAACTCTTTGAGGAAGCACACATC 3235
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTGGAAATACATTTCTTCCACCAGCAATGATGATTTTCTCTATGCTGGATGTTTC 3295
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 3296 CGGTTTGGAGCCTACTTGTGGCACAATAACTCATGAGCTTTGAGGATGTTCTGTAGTA 3355
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProasp 999
Db 3356 TTTTACGCTGTGCTTTTGGTGCCATGCGCGGCGGCAAGTCAGTTCATTTGCTCTCTGAC 3415
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAAGCCAAATATACAGAGCCCATCATCATGATCATTTGAAAAACCCCTTTG 3475
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTGACAGCTACAGCAGCAAGGCGCTTAATGCCGACACATTTGGAAGAAATGTCACATTT 3535
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGATTAACATATCCACCACCGGACATCCCACTGCTTCAGGAGCTGAGC 3595
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3596 CTGAGGTGAAGAAGGCCAGACGCTGGCTCTGGTGGGACAGTGGCTGTGGGAAGAC 3655
QY 1080 ThrValValGlnLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3656 ACAGTGGTCCAGCTCTGGAGCGGTTCTACGACCCCTTGGCAGGAAAGTGTCTGTAT 3715
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 GGCAGAAGAATAAGCGACTGAATGTTCAAGTGGCTCCGAGACACCTGGGCACTGTGCC 3775
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 CAGGAGCCCATCTGTTGACTGACGATTTGAGAACATTTGCCATGATGAGACACACAGC 3835
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisPhe 1159
Db 3836 CGGGTGGTGTCAAGGAAGAGATCGTGGGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGG 3895
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCGAGTCACTGCTTAATAATATAGCACTAAAGTAGGAGACAAAGAACTCAGCTCTCT 3955
QY 1180 GlyCylGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GGTGGCCAGAAACACGCAATGTCATAGCTGCTGCGCTTTGTAGACAGGCTCATATTTG 4015
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGGATGAGCCAGCTGCTGAGTCTGGATACAGAAGTGAAAGGTTGTCCAAAGAGCC 4075
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGACAAAGCCAGAGAGGCCGACCTGCATTTGATTGCTCACCGCTGTCCACCATC 4135
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGAATGCACACTTAATAGTGGTGTTCAGAAATGGCAGATCAAGGAGGATGGCAGCAT 4195
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGTGGCAGAAAGGCATCTATTTTCAATGGTCACTGCTCCAGGCTGGAACA 4255

QY 1280 LysArg 1281
Db 4256 AAGGC 4261
RESULT 6
US-08-784-649A-1
; Sequence 1, Application US/08784649A
; Patent No. 5830697
; GENERAL INFORMATION:
; APPLICANT: Sikic, Branimir I
; APPLICANT: Chen, Gang
; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
; CYCLOSPORIN MODULATION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/784,649A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: Reg.No. 5830697 36,677
; REFERENCE/DOCKET NUMBER: 06037/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-322-5070
; TELEFAX: 415-854-0875
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4264 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-784-649A-1
Alignment Scores:
Pred No.: 0 Length: 4264
Score: 5815, 50 Matches: 1162
Percent Similarity: 94.93% Conservative: 55
Best Local Similarity: 90.64% Mismatches: 60
Query Match: 89.84% Indels: 6
DB: 2 Gaps: 5
US-09-672-725C-25 (1-1281) x US-08-784-649A-1 (1-4264)
QY 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 143 ATGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAAAGAAAGAACTTTTAAACTGAAC 202
QY 20 LysLysSerLysLysGluLysLysLysLysProThrValSerThrPheAlaMet 39
Db 203 AATAAAGT---GAAAAAGATAAGAAAGAAAGAAAGAACTGTCAGTGTATTTCAATG 259
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
Db 260 TTTCCATTCAAAATG-CTTGACAAAGTTGATATGTTGGTGGGAACTTTGGCTGCCATC 318
QY 60 IleHisGlyAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 319 ATTCATGGGCTGGACTTCCTCTCATGATGCTGTTGGAGAAATGACAGATATCTTT 378
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99

[illegible]

```
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
|||||
Db 2590 CTCGCCAATGATGCTGCTCAAGTAAAGGGGCTATAGGTTCAGGCTTCGTAAATTACC 2649
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleIleIleIleIle 859
|||||
Db 2650 CAGAATATAGCAAACTTGGGACAGGAATAATATATATCTCTCATCTATGTTGGCAACTA 2709
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
|||||
Db 2710 ACACGTCTTACTCTTACCAATTTGACCATTCATGCAATAGCAGGAGTGTGTGAATGAAA 2769
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 899
|||||
Db 2770 ATGTTCTCTGGCAACACCTGAAGATAGAAGTGTCTGGGAAGATCGCT 2829
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
|||||
Db 2830 ACTGAAGCAATAGAAAACCTTCGAACCGTGTGTTCTTGTGACTCAGGAGCAGAAAGTTTGA 2889
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
|||||
Db 2890 CATATGTATGCTCAGAGTTTGGAGGTACCACATACAGAACTCTTTGAGGAACACACATC 2949
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 2950 TTTTGGAAATACATTTTCCCTCACCCAGGCAATGATGATTTTCTTCTATGCTGATGTTTC 3009
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
|||||
Db 3010 CGGTTTGGAGCCCTACCTTGGTGACACATACTCATGAGCTTTGAGGATGTTCTGTAGTA 3069
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProasp 999
|||||
Db 3070 TTTTCAGCTGTGCTGTGTGTCATGCGCGGTGGGCGCAAGTCAGTTCATTTCTCTCTGAC 3129
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
|||||
Db 3130 TATGCCAAAGCCAAATATATCAGCAGCCCATCATCATCATCATGATCATTTGAAAACCCCTTG 3189
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
|||||
Db 3190 ATTGCAGCTTACAGCAGCGGAGCCCTTAATGCCGACACATTTGGAAGAAATGTCACTTT 3249
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
|||||
Db 3250 GGTGAAGTTGTATTAACCTATCCACCCGACCCGACATCCAGTGTCTCAGGACTGAGC 3309
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValIleGlySerSerGlyCysGlyLysSer 1079
|||||
Db 3310 CTGGAGGTGAAGAAGGCGCCAGACGCTGGCTCTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3369
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
|||||
Db 3370 ACATGTGTCAGCTCTCGGAGCGGTCTACGACCCCTTGGCAGGAAAGTGTGCTGTAT 3429
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTyrLeuArgAlaHisLeuGlyIleValSer 1119
|||||
Db 3430 GGCNAAAGAAATAAGCCACTGATGTTTCACTGCTCCGAGCACACCTGGCATCGTGTCC 3489
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
|||||
Db 3490 CAGGAGCCCATCTGTTTGTACTCAGCATTTGCTGAGAACAATTTGCTATGGAGACAACAGC 3549
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
|||||
Db 3550 CGGTTGTGTACAGGAAGAGATTTGAGGGCGAGCAAGAGGCCCAACATACATGCTCTC 3609
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
|||||
Db 3610 ATCAGGTCACTGCCTAATAATATAGCACTAAAGTAGGAGACAAGAACTCAGCTCTCTCT 3669
```

```
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
|||||
Db 3670 GGTGCCAGAAACAACCATTCATAGCTCGTGCCCTTTAGACAGCCCTCATATTTTG 3729
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
|||||
Db 3730 CTTTGGATGAGCCAGCTGCTGATACAGAAAGTGAAGAGTTGTCCAAAGAGCC 3789
QY 1220 LeuAspLysAlaArgGluArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 3790 CTGCACAAAGCCAGAGAAGCGCCACCTGCATTGTGATTGCTACCCGCTGTCCACCATC 3849
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||
Db 3850 CAGAATCGAGACTTAATAGTGTGTTTTCAGAAATGGCAGAGTCAAGGAGCATGGCAGCAT 3909
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 3910 CAGCAGCTGCTGGCACAGAAAGGATCTATTTTCATGTGTCAGTGTCCAGGCTGGAACA 3969
QY 1280 LysArg 1281
|||||
Db 3970 AAGCGC 3975

RESULT 7
US-08-784-649A-5
; Sequence 5, Application US/08784649A
; Patent No. 5830697
; GENERAL INFORMATION:
; APPLICANT: Sikic, Branimir I
; APPLICANT: Chen, Gang
; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
; TITLE OF INVENTION: CYCLOSPORIN MODULATION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/784,649A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: Reg.No. 5830697 36,677
; REFERENCE/DOCKET NUMBER: 06037/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-322-5070
; TELEFAX: 415-854-0875
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4264 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-784-649A-5

Alignment Scores:
Pred. No.: 0
Length: 4264
Score: 5815.50
Matches: 1162
Percent Similarity: 94.93%
Conservative: 55
Best Local Similarity: 90.64%
Mismatch: 60
Query Match: 89.84%
Indels: 6
Gaps: 5
DB:
```


|||||
Db 2216 ATCGCTGGTTTCGATGATGAGTCATTTGGAGAAAGGAAATCATGATCAACTCATGAAA 2275
Qy 620 GluLysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 2276 GAGAAAGGCATTACCTTCAAACTGTGCACAAATGCACAGCAGGAAATGAAGTTGAATTA 2335
Qy 640 GluAsnAlaThrGlyGlySerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2336 GAAATGCAGCTGATGATCAACCAAGTGAATTTGATGCCTTGGAAATGCTTCAAAATGAT 2395
Qy 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2396 TCAAGATCATCGTCTAATAAGAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAACCC 2455
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2456 CAAGACAGAAGCTTAGTACCAAGAGGCTCTGGATGGAAGATATACCTCCAGTTTCCTTT 2515
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2516 TGGAGGATTATGAAGCTAAATTTAACTGAATGCCTTATTTGTTGTTGGTGTATTGTT 2575
Qy 720 AlaIleLeasnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2576 GCCATTATAAATGGAGGCTGCACCAAGCAGATTTCGAATAATATTTTCAAAAGATTATAGG 2635
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 GTTTTACAAAGATTGATGATCTCGAACAAGACGACAGAACTAACTTGTGTTTTCACATA 2695
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2696 TTGTTTCTAGCCCTTGGAAATTTATTTCTTTTATATACATTTTCTTCAGGGTTTCACATTT 2755
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 GGCAAAAGCTGGAGAGATCCTCACAAAGCGCTCCGATACATGTTTCCGATCCATGCTC 2815
Qy 800 ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2816 AGACAGATGTGAGTTGGTTGTATGACCTTAAACACACACTGGAGCATTGACTACCAGG 2875
Qy 820 LeuAlaAsnAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2876 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTCTGCTGAATTACC 2935
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2936 CAGAAATAGCAATCTTGGGACAGGAATAATATATCTTCTATGTTGGCAACTA 2995
Qy 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2996 AACTGTTACTCTTAGCAATTTGACCCATCATGTCAATAGCAGGAGTGTTCGAATGAAA 3055
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluAlaGlyLysIleAla 899
Db 3056 ATGTGTGCTGGCAAGCACTGAAAGATAAGAAAGAACTAGAAGGTCTGGGAAGATCGCT 3115
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 3116 ACTGAAGCAATAGAAACTTCCGAACCGTTGTTCTTCTTCTACGAGCAGAGAGTTTGA 3175
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGTATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAACACACATC 3235
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTTGAATTACATTTTCTTCCACCAAGCAATGATGATATTTTCTATGCTGGATGTTT 3295
Qy 960 ArgPheGlyAlaTyrIleuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
|||||

Db 3296 CGGTTTGAGCGCTACTTGGTGGCACATAAACTAATAGCTTTTGAGGATGTTCTTTAGTA 3355
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3356 TTTTCAGCTGTGTCTTGGTGCCATGGCCGTGGGCAAGTAGTTCATTTCCTCTGAC 3415
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAAGCCAAAATATCAGCAGCCACATCATCATGATCATTTGAAAAAACCCCTTG 3475
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTGACAGCTACAGCAGGAAGCCCTAATGCCGAACACATTTGGAAGGAATGTCCACATTT 3535
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGTATTCAACTATCCACCCGCGCAGCATCCAGTCTTCAGGGACTCAGC 3595
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3596 CTGGAGGTGAAGAGGCCAGACGCTGGCTCTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3655
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValIleAsp 1099
Db 3656 ACAGTGTGTCAGCTCCCTGGAGCGGTTCTACACCCCTTGGCAGGGAAGTCTCTGTAT 3715
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 GGCAAGAAATAAAGCGACTGAATGTTTCAGTGGCTCCGAGCAGACCTGGGCATCGTGCC 3775
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 CAGGAGCCCATCTCTGTTGACTGCAGCATTTGCTGAGAACATTTGCTATGGACACACAGC 3835
Qy 1140 ArgValValSerHisGluGluIleMetClnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3836 CGGTTGGTGTCCAGGAAGAGATCGTAGGGCAGCAAGGAGGCCACATACATGCGCTTC 3895
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCGAGTCACCTGCCTTAATAATATAGCACTAAAGTAGGAGACAAAGGAACCTCAGCTCTCT 3955
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GGTGCCAGAAACCAACGCAATTCGATGCTGCTGCCCTTGTAGACAGCCTCATATTTTG 4015
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGGATGAAGCCACGCTCAGCTCTGGATACAGAAAGTGAAAAGGTTGTCAGAGAGCC 4075
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGGACAAACCCAGAGAGGCCGACCTGCATTTGTTGTTGCTACCGCTGTCCACCATC 4135
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGAATGCAGACTTAATAGTGTGTTTTCAGAAATGCGCAGAGTCAAGAGCATGGCACCAT 4195
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGCTGGCACAGAAAGGCAATCTATTTTTTCAATGGTCAGTGTCCAGCCTGGACA 4255
Qy 1280 LysArg 1281
Db 4256 AAGCGC 4261

RESULT 9
US-09-120-513-1
; Sequence 1, Application US/09120513
; Patent No. 6025160
; GENERAL INFORMATION:
; APPLICANT: Brun, Kimberly
; APPLICANT: Chenerly, Richard


```
Db 3608 CTTCTGGATGAAGCGACATCAGCTCTGGATACGGAGAGTGAAGAGGTGCTCCAGGAAGCG 3667
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3668 CTGGACAAGCCAGGAAGCGCCACCGCTTGTGATCGCGCACCCGCTCTCCACCATC 3727
QY 1240 GlnAsnAlaAspLeuIleValIlePheGlnAsnGlyLysValIysGluHisGlyThrHis 1259
Db 3728 CAGAACGCAGACTTGATCGTGGTATTACAGAACGCCAGGTCAGAGGACGCGGCCACCAC 3787
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIlePheSerMetValSerValGlnAlaGlyAla 1279
Db 3788 CAGCAGCTGCTGGCCAGAAAGGATCTATTCTCGATG-----GTTCAGGCTGGAGCA 3841
QY 1280 LysArg 1281
Db 3842 AAGCGC 3847

RESULT 10
US-09-450-105-1
; Sequence 1, Application US/09450105
; Patent No. 6169166
; GENERAL INFORMATION:
; APPLICANT: Kimberly Anne Brun
; APPLICANT: Richard James Chenery
; APPLICANT: Harma Ellens
; APPLICANT: John Anthony Felid
; APPLICANT: Lin Yue
; TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE SEQUENCES
; FILE OF INVENTION: ENCODING RAT MDR1B2 AND SCREENING METHODS THEREOF
; FILE REFERENCE: GP-50008-D1
; CURRENT APPLICATION NUMBER: US/09/450,105
; EARLIER FILING DATE: 1999-11-29
; EARLIER FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4233
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-450-105-1

Alignment Scores:
Pred. No.: 0 Length: 4233
Score: 5291.50 Matches: 1030
Percent Similarity: 90.48% Conservative: 130
Best Local Similarity: 80.34% Mismatches: 113
Query Match: 81.75% Indels: 9
Db: 4 Gaps: 5

US-09-672-725C-25 (1-1281) x US-09-450-105-1 (1-4233)
QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 26 ATGGAGTTTGAAGAGGCGCTTAACGGAAACACGACAGCAAGAACTCTCAAAAGATGGGCAA 85
QY 21 LysSerLysLysGluLysLysGluLysGluLysProThrValSerThrPheAlaMetPhe 40
Db 86 AAGAGTAAAG-----GAGAAGGAGAAGAAACCTGCTGTGGCATATTTCGGGATGTTT 139
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 60
Db 140 CGCTATGCAGATTGGCTTGACAAGCTGTGCATGCTCTGGAACTCTCGCTGCTATCATC 199
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
Db 200 CACGGAACCCCTGCTCCCTCCCTGATGCTGTGTTCCGATACATGACAGATAGTTTACC 259
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 260 CAAGCA-----GAGACCCGCATTCTCCGAGCGTTACTTAATCAAGATGAATCAAC 310
```

```
QY 101 AsnThrGlnHisPheIle---AsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 311 AGTACACAGACCGTCAGCGACAGCAGTCTGGAGGAGACATGCCCATGTACGCGCTACTAT 370
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 371 TACACGGGCATTGGTCCGGTGTCTCATCGTTGCCGTACATCCAGCTTTCACCTTGGTGC 430
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 431 CTGGCAGCTGGGAGACAAATACACAAGATTAGCACAAGTTTTTCCATGCCATCATGAAT 490
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 491 CAGGAGATAGGCTGGTTTGACGTGAATGACGTGGGAGCTCAACACCGGCTCACAGAT 550
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db 551 GACGCTCTCCAAAATTAAATGACGGAATTGGTGACAAAATTGGAATGTTCTTCAGTCCATA 610
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrVal 219
Db 611 ACACATTTTTCACCGGTTTTTATAATAGGATTTATAAGTGGTGGAAAGTAACCTTGTGA 670
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 671 ATTTTGGCGTCAGCCCTCTTATTTGGTGTCTATCTGCCATGGCCAAGGTACTGACT 730
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluIleVal 259
Db 731 TCATTACTTAATAAGGAACCTCCAGGCTTATGCGAAAGCTGGAGCAGTCCCGCAAGAAGTC 790
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 791 TTAGCAGCCATCAGAACTGTGATTGCGTTGGAGGACAAAGAAAGAACTTCAAAAGGTAC 850
QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 851 AATAAAATTTAGAGAAGCTAAAGAGTTGGCATAAAGAAAGCCATCAGGCCCAACATT 910
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 911 TCCATAGGTATTGCCCTACCTGTTGGTCTATGCTCTTATGCACACTGTCATGCTGGTATGG 970
QY 320 ThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSer 339
Db 971 ACCTCCTTGGTCTCTCAAAATGAATATTCTATTGGACAAAGTCTTACCGCTCTTCTCTCT 1030
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1031 ATTTTATTGGGACITTCAGTATTGGACATTTAGCCCCAAACATAGAGCCCTTTCGCAAT 1090
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1091 GCAAGAGGGCGACGCTATGAAATCTTCAAGATAATTGATAATAGCAAGCATTGCACAGC 1150
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1151 TTCTCAACCAAGGGACACAAACACAGACAGATAAATGGGAAATTTGGAAATTTAAAAATGTT 1210
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1211 TACTTCAACTACCCATCAGGAAGTGAAGTTAAGATCTTGAAGGGCCTCAACCTGAAGGTG 1270
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1271 AAGAGCGGCAGACGCTAGCCCTGGTTGGCAACAGTGGCTGTGGAAAAAGCAACACTGTC 1330
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1331 CAGCTGCTCAGAGGCTCTTACGACCCCATAGAGGGCGAGGTACGATGATCAGCGACAGAC 1390
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
```


QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
 DB 3608 CTTCTGGATGAGGACATACAGCTCTGGATACGAGAGTGAAGAGTCGTCAGGAAGCG 3667
 QY 1220 LeuAspLysAlaArgGluGlyArgThrCysTyleValIleAlaHisArgLeuSerThrIle 1239
 DB 3668 CTGGACAAAGCCAGGAAGCGCCACCTGCTGATCGCGCACCGCTGTCCACCATC 3727
 QY 1240 GlnAsnAlaAspLeuIleValIleValIlePheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
 DB 3728 CAGAACCGAGACTTGATGCTGGTGAATCAGAACCGCCAGGTCAAGGAGCACGGCACCCAC 3787
 QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
 DB 3798 CAGCAGCTGCTGGCCCGACAAAGGCATCTATTTCGATG-----GTTCCAGCTGGAGCA 3841
 QY 1280 LysArg 1281
 DB 3842 AAGCGC 3847

RESULT 11
 US-08-461-823-1
 ; Sequence 1, Application US/08461823
 ; Patent No. 5593840
 ; GENERAL INFORMATION:
 ; APPLICANT: Bhatnagar, Satish K.
 ; APPLICANT: George Jr., Albert L.
 ; APPLICANT: Nazarenko, Irina
 ; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACID SEQUENCES
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OncorPharm, Inc.
 ; STREET: 200 Perry Parkway
 ; CITY: Gaithersburg
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20877

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/461,823
 FILING DATE: 05-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/168,621
 FILING DATE: 16-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/010,433
 FILING DATE: 27-JAN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Karta, Glenn E.
 REGISTRATION NUMBER: 30,649
 REFERENCE/DOCKET NUMBER: PA-0012 CIP 2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301 527-2058
 TELEFAX: 301 208-6997
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2726 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-461-823-1

Alignment Scores: 0 Length: 2726
 Pred. No.:

Score: 3628.00 Matches: 716
 Percent Similarity: 96.28% Conservative: 35
 Best Local Similarity: 91.79% Mismatches: 29
 Query Match: 56.05% Indels: 0
 DB: 1 Gaps: 0
 US-09-672-725c-25 (1-1281) x US-08-461-823-1 (1-2726)
 QY 502 IleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsnLys 521
 DB 2 ATTGAGAAAGCTGCAAGGAAGCCAAATGCTATGACTTTATCATGAAGACTGCCTCATAAA 61
 QY 522 PheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArgIle 541
 DB 62 TTTGACACCTGTTGGAGAGAGAGGGGCCACGTTGAGTGGTGGGCGAGAGAGATC 121
 QY 542 AlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThrSer 561
 DB 122 GCCATTGCCAGTCCCTGCTTCCAAACCCCAAGATCCTCTGCTGGATGAGCCAGGTCA 181
 QY 562 AlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArgLysGly 581
 DB 182 GCCTTGGACACAGAAAGCGAAGCAGTGGTTCAGTGGCTCTGGATGAAGCCAGAAAGGT 241
 QY 582 ArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla 601
 DB 242 CGGACCAACCATTTGATAGCTCATCGTTTCTCTACAGTTTCGTAATCTCACGTCATCGCT 301
 QY 602 GlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGluLys 621
 DB 302 GGTTCGATGATGGAGTCATTGTGGAGAAAGAAATCATGATGAATCATGAAGAGAAA 361
 QY 622 GlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluLeuGluAsn 641
 DB 362 GGCATTACTTCAAACTGTGCACATGCAGACAGAGAAATGAAGTGAATAGAAAAT 421
 QY 642 AlaThrGlyGluSerLysSerGluSerAlaLeuGluMetSerProLysAspSerGly 661
 DB 422 GCAGCTGATGAATCCAAAGTGAATGATGCCTTGGAAATGCTTCAAAATGATTTCAAGA 481
 QY 662 SerSerLeuLysArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAsp 681
 DB 482 TCCAGTCTAATAAGAAAGATCAACTCGTAGGAGTGTCCGTGGATCAGAACGCCAAGAC 541
 QY 682 ArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheThrArg 701
 DB 542 AGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCTTTTGGAGG 601
 QY 702 IleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAlaIle 721
 DB 602 ATTATGAAGCTAAATTTAACTGAATGGCCTATTATTTGTTGTTGTTGTTGTTGTTGTT 661
 QY 722 IleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePhe 741
 DB 662 ATAAATGGAGGCTGCAACCCAGCATTTGCAATAATATTTTCAAGATTTATAGGGGTTTT 721
 QY 742 ThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe 761
 DB 722 ACAAGAATTGATGATCTGAAACAAACAGACAGAAATAGTAACCTGTTTTCACATTGTTT 781
 QY 762 LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys 781
 DB 782 CTAGCCCTTGGAAATATTCTTTTATTATACATTTTCTCAGGGTTTCACATTTGGCANA 841
 QY 782 AlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGln 801
 DB 842 GCTGGAGAGATCTCCCAAGCGGCTCCGATACATGTTTCCGATCCATGCTCAGACAG 901
 QY 802 AspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla 821
 DB 902 GATGTGAGTTGGTTGATGACCTCAAAACACCACTGGAGCATTTACTACAGGCTCGCC 961
 QY 822 AsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn 841

```
Db 962 AATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGTGTAATTTACCCAGAAAT 1021
Qy 842 ILeAlaasnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeu 861
Db 1022 ATAGCAAACTCTGGACAGGAATAATATATCTCTCATCTATGTTGGCACTAACACTG 1081
Qy 862 LeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMetLeu 881
Db 1082 TTACTCTTAGCAATTTGACCATCTGCAATAGCAGGAGTTGTTCAAAATGAAATGTTG 1141
Qy 882 SerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAlaThrGlu 901
Db 1142 TCTGGCAACAGCTAGAAAGATAGAAAGAACTAGAGGTGCTGGGAAGATCGCTACTGAA 1201
Qy 902 AlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyrMet 921
Db 1202 GCATAGAAAACCTCTCGAACCGTTGTTCTTGTACTCAGGACGACAAAGTTTGAACATAG 1261
Qy 922 TyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGly 941
Db 1262 TATGCTCAGAGTTGTCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTTTGA 1321
Qy 942 ValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPhe 961
Db 1322 ATTACATTTTCTTACCAGGCAATGATGATATTTTCTATGCTGGATGTTTCCGGTTT 1381
Qy 962 GlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSer 981
Db 1382 GGAGCCTACTTGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTATTTC 1441
Qy 982 AlaIleValPheGlyAlaMetAlaValGlyGlnValSerPheAlaProAspTyrAla 1001
Db 1442 GCTGTTGCTCTTGGTGCCATGCGCGTGGGCAAGTCAGTTTCATTGCTCCTGACTATGCC 1501
Qy 1002 LysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIleAsp 1021
Db 1502 AAAGCCAAATATCACAGCCACCATCATCATGATCATTTGAAAACCCCTTTGATTGAC 1561
Qy 1022 SerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyYasnValThrPheAsnGlu 1041
Db 1562 AGCTACAGCAGGAGGCTTAATGCCGACACATTTGGAAGGAAATGTCACATTTGGTGA 1621
Qy 1042 ValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeuGlu 1061
Db 1622 GTTGATTTCAACTATCCACCGCGGACATCCAGTCTTCAGGAGCTGCTGCTGATGGCAA 1681
Qy 1062 ValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrVal 1081
Db 1682 GTGAAGAGGGCCAGCGCTGCTGTGGTGGCAGCGTGGCTGTGGGAAGAGCACAGTG 1741
Qy 1082 ValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGlyLys 1101
Db 1742 GTCCAGCTCCTGAGCGGTTCTACGACCCCTTGGCAGGGAAGTGTCTGTTGATGGCAA 1801
Qy 1102 GluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyYileValSerGlnGlu 1121
Db 1802 GAAATAAGCGACTGAATGTTTCAGTGGCTCCGAGCACACCTGGGCATCGTGTCCAGGAG 1861
Qy 1122 ProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgVal 1141
Db 1862 CCCATCCTGTTGACGTCAGCATTTGCTGAGAATGTTGCTATGGAGAACACAGCCGGGTG 1921
Qy 1142 ValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisPheIleGlu 1161
Db 1922 GTGTACAGGAAGAGATCGTAGGGCAGCAAGAGGCAACATACATGCTTCATCGAG 1981
Qy 1162 ThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGly 1181
Db 1982 TCACCTCCCTTAATAATATAGCACTATGAGTATGAGGAGCAAGGAACTCAGCTCTCTGGTGC 2041
Qy 1182 GlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeu 1201
```

```
Db 2042 CAGAAACAACGCATTGCCATAGCTCGTCCCTTGTGTAGACAGCCTCATATTTGCTTTTG 2101
Qy 1202 AspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeuAsp 1221
Db 2102 GATGAAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAACCCCTGGAC 2161
Qy 1222 LysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsn 1241
Db 2162 AAAGCCAGAGAGCCGACCTGCATTTGTGATTGCTCACCGCTGTCCACCATCCGAAT 2221
Qy 1242 AlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGlnGln 1261
Db 2222 GCAGACTTAATAGTGGTGTTCAGAAATGGCAGATCAAGAGCATGGCAGCATCAGCAG 2281
Qy 1262 LeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLysArg 1281
Db 2282 CTGCTGGCACAGAAAGCATCTATTTTCAATGTCAGTGTCCAGGCTGGAACAAGCGC 2341

RESULT 12
US-08-612-734B-1
; Sequence 1, Application US/08612734B
; Patent No. 5914246
; GENERAL INFORMATION:
; APPLICANT: Peery, Robert B.
; APPLICANT: Skatrud, Paul L.
; APPLICANT: Tobin, Matthew B.
; TITLE OF INVENTION: Multiple Drug Resistance Gene of
; TITLE OF INVENTION: Aspergillus Fumigatus
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center, DC1501
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.A.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,734B
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Craig, Anne I.
; REGISTRATION NUMBER: 32,976
; REFERENCE/DOCKET NUMBER: X-9681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-354-9570
; TELEFAX: 617-354-4043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4047 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4047
US-08-612-734B-1

Alignment Scores:
Pred. No.: 1,13e-275 Length: 4047
Score: 2513.00 Matches: 559
Percent Similarity: 59.43% Conservative: 229
Best Local Similarity: 42.16% Mismatches: 464
Query Match: 38.82% Indels: 74
DB: 2 Gaps: 17

US-09-672-725C-25 (1-1281) x US-08-612-734B-1 (1-4047)
```

QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLys----- 13
DB 148 ATGGCGCCTCAGATGGCAAGAAAGATCACGGAAAGCTGTCGACCTGAACGATGAT 207
QY 14 -----AsnPheTrpLysMetGlyLysLysSerLysLysLysGluLysGlu 29
DB 208 TCTCTCTTTGGCCATCTCCAGGACGACGAGAAAGAGGTACTCAAGCGACAGCTTGATGCA 267
QY 30 LysLysProThrValSerThrPheAlaMetPheArgTrpSerAsnTrpLeuAspArgLeu 49
DB 268 CCGTCTGTGAAGTCTCTTCTTCACTTGTATCGCTATGCTCGCTCGAGAAAGACATCTG 327
QY 50 TyrMetLeuValGlyThrMetAlaAlaIleIleHisGlyAlaAlaLeuProLeuMetMet 69
DB 328 ATCATATCTGTCGCGCAATCTGTCTATCTGCTGCTGGTGGCTCTTCCCTGTTTCA 387
QY 70 LeuValPheGlyAsnMetThrAspSerPheAlaAsnAlaGlyIleSerArgAsnLysThr 89
DB 388 ATTCTCTTCGGTTCACATAGCCTCAGCGTTC-----CAGGCAATATCTCTGGGT---ACT 438
QY 90 PheProValIleIleAsnGluSerIleThrAsnAsnThrGlnHisPheIleAsnHisLeu 109
DB 439 ATGCCC-----TATCATGAGTTCTATCAC--- 462
QY 110 GluGluGluMetThrThrTyrAlaTyrTyrSerGlyIleGlyAlaGlyValLeuVal 129
DB 463 -----AAACGTACTAGAAATGCTTACTTGTGTATCTCGGTATTCGGAGTTGTC 516
QY 130 AlaAlaTyrIleGlnValSerPheTrpCysLeuAlaAlaGlyArgGlnIleLeuLysIle 149
DB 517 ACAGTCTATGTCAGCACCGTGGTTCATTATATCTGGCGAAACATCTCACACAGAATC 576
QY 150 ArgLysGlnPhePheHisAlaIleMetArgGlnGluIleGlyTrpPheAspValHisAsp 169
DB 577 CGTGAATAATATCTTGAGGTATCTCGGCAAGTATGCTTACTTTCGACAAAGTTGGGC 636
QY 170 ValGluLeuAsnThrArgLeuThrAspValSerLysIleAsnGluGlyIleGly 189
DB 637 CCGGTGAAGTTACCACGCGTATCATCTGCTGATACCAACCTGATCCAGAGCGCATCTCT 696
QY 190 AspLysIleGlyMetPhePheGlnSerIleAlaThrPhePheThrGlyPheIleValGly 209
DB 697 GAGAAAGTTGCTTCACTTTGACCGCATTCGCCACATTTGTAACCCGCAATTTATTGTCGCC 756
QY 210 PheThrArgGlyTrpLysLeuThrLeuVal-----IleLeuAlaIleSerPro 225
DB 757 TACGTCAAGTATTTGGAAGTTGGTCTGTATCTGTACCTCAACTATCTGTCGCGCTGGTCATG 816
QY 226 ValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPheThrAspLysGlu 245
DB 817 GTTATGGAGGTGG-----TCGAGGTTTATGTGAAGTHACACAGAATCT 864
QY 246 LeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeuAlaAlaIleArgThr 265
DB 865 ATTTAAAGTTATGTGTGTGGAACTTCGCGGAAGAGTCAATCAGCTCCATTCGGAAT 924
QY 266 ValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTrpAsnLysAsnLeuGluGlu 285
DB 925 GCTACCGCTTTCGCACTCAGGATAAGCTGCCAAGCAATACGAAACCCATCTGGCTGAG 984
QY 286 AlalysGlyIleGlyIleLysAlaIleThrAlaAsnIleSerIleGlyAlaAlaPhe 305
DB 985 GCTGAAAATGGGCGTCAACACACAGGTATCTCTGGTATGATGATGTGGTATGTTCT 1044
QY 306 LeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTrpGlyThrSerLeuValLeuSer 325
DB 1045 GGTATCATGTTCTCGAACTATGGTCTCGGTTCTTGGATGGATCATCTCGTTCGTTGCGGT 1104
QY 326 SerGluTyrSerIleGlyGlnValLeuThrValPhePheSerValLeuIleGlyAlaPhe 345
DB 1105 AAAGAAGTCAACGTGGGCCAAGTTCTGACAGTTTTGTATGTTGATCTATCTGATCGGTGCTTC 1164

QY 346 SerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArgGlyAlaAlaTyr 365
DB 1165 AGTTTGGCAACGTGCCCCCAATGGTGTAGCCCTTTACGAATGGTGTCTGCTGGCGCGG 1224
QY 366 GluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSerLysSerGlyHis 385
DB 1225 AAGATTATACAGCAGATTGACCCAGATCGCCACTGGACCCCTATTCTGACGAAGGAAG 1284
QY 386 LysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPheSerTyrProSer 405
DB 1285 GTACTCGACCATTTTGAAGAAATATCGAATTTGCAATGTCAACACATCATCCCTTCA 1344
QY 406 ArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSerGlyGlnThrVal 425
DB 1345 AGACCCGAAGTTACAGTCATGGAAGATGCTCTTTATCATGCCCCCGGAAGATAC 1404
QY 426 AlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeuMetGlnArgLeu 445
DB 1405 GCATTGGTGGGCCATCTGGCTCTGGAAGAGTACTGTTGCGGCTTGGTGAGCGCTTT 1464
QY 446 TyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArgThrIleAsnVal 465
DB 1465 TACCTTCCAGTAGGAGGCCAGGTATTGCTGGACGGCCATGATATCCAAACCTCAACCTC 1524
QY 466 ArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAlaThrThr 485
DB 1525 CGTGTGCGACAGCAGATCTCTCTGTGACGCCAGAAACCTGTCTTTTCAGCACCAG 1584
QY 486 IleAlaGluAsnIleArgTyrGlyArg-----GluAsnValThrMetAsp 500
DB 1585 ATCTTTAGAAACATCAACATGCTTGTATTGGCCACCAAAATTCAGCATCAGTCGAAGGAC 1644
QY 501 GluIle-----GluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMet 516
DB 1645 AAGATTAGAGAGCTCCTGTGAGATGCGGCAGAAATGCGCAATGCTCATGATTTATTATG 1704
QY 517 LysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGly 536
DB 1705 GCTCTGCTTGAAGTTTACGATAGCAATGAGTGGGTACGCGTGTCTTACTTTCAGGAGT 1764
QY 537 GlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeu 556
DB 1765 CAGAAACAAGTATTGCCATTTGCTGTCGCAATGTCAGTGACCCCAAGATTTCTGTGCTT 1824
QY 557 AspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAsp 576
DB 1825 GATGAAGCTACATCAGCTTTGGATACCAACGAGGCGTCTGCCAAGCCCTCTTGAT 1884
QY 577 LysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsn 596
DB 1885 AAAGCTGCGGAGGTAGAATACCATTTGCTACCTGTTGCTACCAATCAAAACA 1944
QY 597 AlaAspValIleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGlu 616
DB 1945 GCCCAACAATTTGTCCTTCCATGTCGCGCAAGATTGCCGAACAGGGAACACATGACGAA 2004
QY 617 LeuMetLysGluLysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGlu 636
DB 2005 TTGGTCGATCGCAAAAGGACGCTACTATAACTTGTGGAGCGCAGCTATCAACGAGGAG 2064
QY 637 IleGluLeuGluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSer 656
DB 2065 AAGGAACAGAA-----GCTCTGGAAGCCGACGCC---CACATGGAC 2103
QY 657 ProLysAspSerGlySerSerLeuIleLysArg-----ArgSerThrArgArg 672
DB 2104 GCGGATGATTTGGTCAAGAAGGGTTACTCGCATCAAGACTGCAGTTAGCAGCTCGAAT 2163
QY 673 SerIleHisAlaProGlnGlyGlnAspArg-----LysLeuGlyThrLysGlu 688
DB 2164 TCTCTCGATGCTGTAGATGAGAAGCGCGCTTGGAGATGAAGCGAAGCCGGAACGAAAA 2223
QY 689 Asp-----LeuAsnGluAsnValPro-----ProValSerPhe 699

```
Db 2224 TCGGTTTCAGTCCGCTTCTATCAAGAAAGTCCCGAGCAGTTCGAGAAGTATTCACTC 2283
|||:|||||
Qy 700 TrpArgIleLeuLys-----LeuAsnSerThrGluTrpProTyrPheValVal 715
|||:|||||
Db 2284 TGGACCTCGTCAAGTTCATCGGTGCTATTAACCGCCCTGAGCTGGTTACATGCTCAT 2343
|||:|||||
Qy 716 GlyPheCysAlaIleIleAsnGlyLeuGlnProAlaPheSerIleIlePheSer 735
|||:|||||
Db 2344 GCGTTGACCTTCGCTTCGCGTGTGCTGCTCAACCTACCCAGGCTTTCTGTACGCC 2403
|||:|||||
Qy 736 ArgIleIleGlyPheThrArgAspGluAspProGluThrLys---ArgGlnAsnSer 754
|||:|||||
Db 2404 AAGGCAATTAGCAGCGCTCGTTACCGAATCAATGTTTCACAAACTCAGGCATGACGCG 2463
|||:|||||
Qy 755 AsnMetPheSerValLeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeu 774
|||:|||||
Db 2464 AATTTCTGCTGCTGATGTTCTGCTGTTGGAATTTGCTCAATTTATCAGCCTGTCTATC 2523
|||:|||||
Qy 775 GlnGlyPheThrPheGlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetVal 794
|||:|||||
Db 2524 AATGGTACAGCATTTGCTATTGTTCCGAGAGACTCATTCGCGGAGCTAGAAAGTCAAGCG 2583
|||:|||||
Qy 795 PheArgSerMetLeuArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGly 814
|||:|||||
Db 2584 TTTAGATCGATTCTTCGTCAGGACATCTCATTTTCGACAGGAAGAGAACACGACCGGT 2643
|||:|||||
Qy 815 AlaLeuThrThrArgLeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArg 834
|||:|||||
Db 2644 GCCTTGACGCTTCCTTATCAACAGACGAGCAAGATCTCTCGGGCGGTAGCGGAGTGACT 2703
|||:|||||
Qy 835 LeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIle 854
|||:|||||
Db 2704 CTCGGCACATCATCATGACTAGCATCTGCGCGCGCATGATCATTCATTCGCGG 2763
|||:|||||
Qy 855 TyrGlyTrpGlnThrLeuLeuLeuLeuAlaIleValProIleAlaIleAlaGly 874
|||:|||||
Db 2764 ATCGGATGGAAGCTGCTCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 2823
|||:|||||
Qy 875 ValValGluMetLysMetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGly 894
|||:|||||
Db 2824 TTCCTCAGATTTCATGCTGCTGCTCAATTCAGCAAGCATCGAAGTCTGCTTCAGGAGG 2883
|||:|||||
Qy 895 AlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArg 914
|||:|||||
Db 2884 TCTGCGAGCTATGCTTGCAGAGCCAGCTCAGCGATCCGCACTGTAGCATCATCTGCG 2943
|||:|||||
Qy 915 GluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeu 934
|||:|||||
Db 2944 GAACAAGATGCTCGGGCGGTTTACCCAGCACAGCTACAAAAACAGGGAGAGAGTTG 3003
|||:|||||
Qy 935 ArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSer 954
|||:|||||
Db 3004 ATCTCAGTCTGAGATCCCTCCCTGCTGTATGCTGCTGCGCAGCATTTGTTTCTGCG 3063
|||:|||||
Qy 955 TyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGln 974
|||:|||||
Db 3064 GTCGCTTTGGGCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3123
|||:|||||
Qy 975 AspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSer 994
|||:|||||
Db 3124 CGTCTTCTGCTGCTTCTGATGTTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3183
|||:|||||
Qy 995 SerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIle 1014
|||:|||||
Db 3184 TCTTTGCCCCGACATGGGTAGGCAAGAAAGTACGCTGCTCAATTCAGAAACTCTTC 3243
|||:|||||
Qy 1015 GluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGlu 1034
|||:|||||
Db 3244 GACAGCAAGCAACCATCTGACATCTGCTGAGTGGGCGGAGAAAGTTGGAGTCTATGGA 3303
|||:|||||
Qy 1035 GlyAsnValThrPheAsnGluValPheAsnTyrProThrArgProAspIleProVal 1054
|||:|||||
```

```
Db 3304 GCGAAATCGAATTCGCGGAGCTCCACTTTAGGTACCCAAACGCGGCGGAGACGCTGTT 3363
LeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSer 1074
|||:|||||
Db 3364 CTTGAGGACTGAATTTGAGCGTGAACCTGGCAATATACATTTGCCCTTTGTTGGACCCAGT 3423
|||:|||||
Qy 1075 GlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGly 1094
|||:|||||
Db 3424 GGATGCGGTAAAGCACTACGATTGCTCTGTTGAGCGATTTTATGACGCACTTGCTGGA 3483
|||:|||||
Qy 1095 SerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHis 1114
|||:|||||
Db 3484 GGGGTCTCTGTTGACGAAAGGACATTAACCAACTCAATGCTCACTCATCCGCAAGTTTC 3543
|||:|||||
Qy 1115 LeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAla 1134
|||:|||||
Db 3544 CTCCTCCCTGTCACCAAGAACCTACTCTGTATCAGGGTACCCTCAAGGAATAATCTCTG 3603
|||:|||||
Qy 1135 TyrGlyAspAsnSerArgValValSerHisGluGluIleMetGlnAlaLysGluAla 1154
|||:|||||
Db 3604 CTTGAGTTCGATAAGGATGAGTTTCGAGGAGACTTTGATTAAAGTCTGCAAGATGCC 3663
|||:|||||
Qy 1155 AsnIleHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLys 1174
|||:|||||
Db 3664 AACATCTATGATTTCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3723
|||:|||||
Qy 1175 GlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArg 1194
|||:|||||
Db 3724 GGAGCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3783
|||:|||||
Qy 1195 GlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLys 1214
|||:|||||
Db 3784 GACCCCAAGTCTCTCTGATGAAGCCACATCTCTCTGATCCGATCTCGAATCTGAGAAA 3843
|||:|||||
Qy 1215 ValValGlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHis 1234
|||:|||||
Db 3844 GTCGTACAAGTCTGCGTGGATGCTGCGCGCGGGCGGACACGATGCTGTGTTGCCAC 3903
|||:|||||
Qy 1235 ArgLeuSerThrIleGlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLys 1254
|||:|||||
Db 3904 CGGCTGAGCACCATTCAAAAGCTGATATAATTTACGTTGTCGACCAAGCAAGATCGTC 3963
|||:|||||
Qy 1255 GluHisGlyThrHisGlnGlnLeuAlaGlnLysGlyIleTyrPheSerMetValSer 1274
|||:|||||
Db 3964 GAAAGTGAGACGACCAAGTGTGATTTCGAAACAAAGGCGCGTATTACGAGCTGGTCAAT 4023
|||:|||||
Qy 1275 ValGlnAlaGlyAlaLys 1280
|||:|||||
Db 4024 CTGAGAGTCTCGGAAG 4041
|||:|||||
```

RESULT 13

```
US-08-996-545-1
; Sequence 1, Application US/08996545
; Patent No. 5928898
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene atrd of
; TITLE OF INVENTION: Aspergillus nidulans
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/996,545

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Webster, Thomas D.

REGISTRATION NUMBER: 39,872

REFERENCE/DOCKET NUMBER: X-11766

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-3334

TELEFAX: 317-276-2763

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4002 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..4002

US-08-996-545-1

Alignment Scores:

Pred. No.:	4,95e-272	Length:	4002
Score:	2481.00	Matches:	538
Percent Similarity:	59.41%	Conservative:	223
Best Local Similarity:	42.00%	Mismatches:	458
Query Match:	38.33%	Indels:	62
DB:	2	Gaps:	8

US-09-672-725c-25 (1-1281) x US-08-996-545-1 (1-4002)

QY	24	LysLysGluLysGluLysGluLysProThrValSerThrPheAlaMetPheArgTyrSer	43
DB	268	AAGCGCAGCTGGAGGAGATCAAGTAACATCTCTTCGCTCTCTGCGGTATGCA	327
QY	44	AsnTrpLeuAspArgLysMetLeuValGlyThrMetAlaIleIleHisGlyAla	63
DB	328	ACAAAGATGGATATCTATCATGTAATCATAGTACAACTCTGTGCCATTCGTCGCGCTG	387
QY	64	AlaLeuProLeuMetLeuValPheGlyAsnMetThrAspSerPheAlaAsnAlaGly	83
DB	388	ACTTCCAGAGTAATGTTATAT	411
QY	84	IleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsnThrGln	103
DB	412	-----CAA	414
QY	104	HisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyrSerGlyIle	123
DB	415	ATCTCGTACGACGAGTCTATGATGAATTGACCAAGACGTAAGTCTCTGCTATACCTC	474
QY	124	GlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheThrCysLeuAlaAlaGly	143
DB	475	GGTATCGCGGAGTTGTCACGTCTATGTTAGTACTGTGCTGCATCTATACCGGAGAA	534
QY	144	ArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGluIleGly	163
DB	535	CACGCCAGCAGAGATCCGGAGTATACCTTGAGTCTATCTGCGCCAGACATTTGGC	594
QY	164	TrpPheAspValHisAspValGluLeuAsnThrArgLeuThrAspAspValSerLys	183
DB	595	TATTTGATAAATCTCGGTGCGGGGAAGTGACCAACCGGTATACAGCCGATACAAACCTT	654
QY	184	IleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAlaThrPhePhe	203
DB	655	ATCCAGGATGGCATTTTCGGAGAAGGTGCGTCTCACTTTGACTGCCCTGGGACATTCGTG	714
QY	204	ThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIleLeuAlaIle	223

DB	715	ACAGCATTCATTACGCTAGCTCAAAATACATGGAAGTTGGCTCTAATTTCCAGCTCAACA	774
QY	224	SerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPheThrAsp	243
DB	775	ATTGTGGCCCTCGTCTCACCATGGGCGGTGTTCTCAGTTTATCATCAAGATGACAGAAA	834
QY	244	LysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeuAlaAlaIle	263
DB	835	AAGTCGCTTGACAGCTACGGTCAGGCGGCACTGTTGCGGAAGAGTTCATCAGCTCCATC	894
QY	264	ArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuLeuArgTyrAsnLysAsnLeu	283
DB	895	AGAAATGCCACAGCGTTTGGCCCAAGACAAGCTTCCGAGCAGATGAGGTCACCTTA	954
QY	284	GluGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIleSerIleGlyAla	303
DB	955	GACGAGCTGAGAAATGGGGAACAAGACAGATGTCATGGTTCATGATTGGCGCC	1014
QY	304	AlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSerLeuVal	323
DB	1015	ATGTTTGGCCTTATGTACTCGAACTACGGTCTTGGTCTTGGATGGGTCTCGTTTCCCTG	1074
QY	324	LeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerValLeuIleGly	343
DB	1075	GTAGATGGTCAGTCGATGGGTGATATTCACAGTCTCATGGCCATCTTGATCGGA	1134
QY	344	AlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArgGlyAla	363
DB	1135	TCGTTCTCTTGGGAAAGCTTAGTCAAAATGCTCAAGCATTTACAACGCTGCGCGCG	1194
QY	364	AlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSerLysSer	383
DB	1195	GCCGCAAAAGATATTGGACGATCGATCGCCAGCTCCCATTTAGATCCATATTCGAACGA	1254
QY	384	GlyHisLysProAspAsnIleLysGlyAsnLeuGluLeuPheLysAsnValHisPheSerTyr	403
DB	1255	GGGAAGACGCTCCACCATTTTGGAGGCCACATTCAGTTAGGCAATGTCAAGCATATTAC	1314
QY	404	ProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSerGlyGln	423
DB	1315	CCATCTAGACCCGAGGTCAACGTCATGGAGGATGTTCTCTGTCAATGCCGCTGGAAGA	1374
QY	424	ThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeuMetGln	443
DB	1375	ACAAACGCTTTAGTCGCGCCCTCTGCTCTGGAAGAACGACGCTGCTGGCTGGTTGAG	1434
QY	444	ArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArgThrIle	463
DB	1435	CGATTCTACATGCTCTTCGCGGTACGGTTTCTGGATGGCCATGACATCAAGGACCTC	1494
QY	464	AsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAla	483
DB	1495	AATCTCGGTGGCTTCGCCAACAGATCTCTTTGGTTAGCCAGGAGCTGTTCTTTTGGC	1554
QY	484	ThrThrIleAlaGluAsnIleArgTyrGly	498
DB	1555	ACACGATTTATAAGAAATATTAGGCAGGTCTCTATCGGCACAAAGTACGAGATGAATCC	1614
QY	499	MetAspGlu	514
DB	1615	GAGGATAAGTCCGGGAATCATCGAGAACGGGCAAAATGCGCAATGCTCATGACTTT	1674
QY	515	IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSer	534
DB	1675	ATTACTGCTTGCCTCAAGGTTATGAGACCAATGTTGGCAGCGTGGCTTCTCCTTTCA	1734
QY	535	GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu	554
DB	1735	GGTGGCCAGAAACAGCGCATTCGCGCCGCTGCTGTTGTAGTACCCCAAAATCCTG	1794
QY	555	LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla	574

D	b	1795	CTCCTGGATGAAGCTACTTTCGGCCCTTGGACACAAATCCGAAGCGCTGGTTCAAGCAGCT	1855
Q	y	575	LeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrVal	594
D	b	1855	TTGGAGGCGCAGCTGAAGCCGGAAGCTACTATTGTGATCGCTCATCGCCTTTCACGAGT	1914
Q	y	595	ArgAsnAlaAspValIleAlaGlyPheAspSpGlyValIleValGluLysGlyAsnHis	614
D	b	1915	AAACAGCGCCACAACATTGTGCTTCATGCGCAATGGCAAAATGTCTGAACGAAGAACTCAC	1974
Q	y	615	AspGluLeuMetLysGluLysGlyIleTyrPheLysPheValThrMetGlnThrArgGly	634
D	b	1975	GATGAATTGGTTGACCGCGAGCGCCTTATCGCAAACTGTGGAGGCTCAACGTATCAAT	2034
Q	y	635	AsnGluIleGlu-----LeuGluAsnAlaThrGlyGluSerLysSerGluSerAsp	651
D	b	2035	GAACAGAAGAAAGCTGACGCCCTTGGAGGACGCGCAGCTGAGGATCTCACAAGATGCAGAT	2094
Q	y	652	AlaLeuGluMetSerProLysAspSerGlySerSer-----Leu	664
D	b	2095	ATTGCCAAATCAAACTCGTCAGCGCATCATCCGATCTCGACGGAAACCCACAAACC	2154
Q	y	665	IleLysArgArgSerThrArgSerIleHisAlaProGlnGlyAsnArgLysLeu	684
D	b	2155	ATTGACGGCAGCGGCACCCACAAGTCTGTTTCCAGCGCGGATCTTCTTAAAGAACCC--	2211
Q	y	685	GlyThrLysGluAsnGluAsnValProProValSerPheTyrPheArgIleLeuLys	704
D	b	2212	-----CCCGAAACAACCTCCGAATAACTCATTTAGGACGCTGCTCAAA	2253
Q	y	705	-----LeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla	720
D	b	2254	TTTGTGCTTCCTTCAACCGCCCTGAATCCGCTACATGCTCATCGGCTTGTCTTCTCA	2313
Q	y	721	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyLe	740
D	b	2314	GTGTTAGCTGGTGGTCCCAACCCACGCAAGCAGTGTATATGTAAAGCCATCAGCAC	2373
Q	y	741	PheThrArgAspGluAspProGluThrLys---ArgGlnAsnSerAsnMetPheSerVal	759
D	b	2374	CTCTCGCTCCAGAAATACAAATATACGAAGCTTCACATGATGGGATTTCTGGTCATTG	2433
Q	y	760	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	779
D	b	2434	ATGTTTCTTCGTGGTGTGATCATCATAGTTTATACGCGCAGTCAACCAATGTGTGCAAT	2493
Q	y	780	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	799
D	b	2494	GCCGTATGCTCCGAGAGACTTATTGTCGCGCGAAGACGACTGCCTTTCGGACGATACTC	2553
Q	y	800	ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg	819
D	b	2554	CGTCAAGACATTGCTTCTTTGACAGAAGGAAGATAAGCAGCGCGCTCTGACCTCTTTC	2613
Q	y	820	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	839
D	b	2614	CTGTCCACCGACGAAAGCATCTCTCCGGTGTAGCGGTGACTCTAGGCACCATCTTG	2673
Q	y	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	859
D	b	2674	ATGACCTCCAGCACCCTAGGAGCGGCTATCATTTATGCCCTGGCATTTGGGTGGAAATG	2733
Q	y	860	ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	879
D	b	2734	GCCTTAGTTTGTATCTCGGTTGTCCGGTTCCTCGGCATGCGGTTTCTACCGATTCTAT	2793
Q	y	880	MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla	899
D	b	2794	ATGCTACCCAGTTTCAATCACGCTCCAAGCTTGCTTATGAGGATCTGCAAACTTGTCT	2853
Q	y	900	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu	919
D	b	2854	TGCCAGGCTACATCGTCTATPCCGCAGTGTGGCTCATTAACCCGGAAGGATGTCTCG	2913

Qy	920	TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle	939
Db	2914	GAGATTTCACATGCCAGCTTCGACGACAAAGCAGCAGCAGCTAACTCTGCTGTGAGG	2973
Qy	940	PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe	959
Db	2974	TCATCCTGTTATATCGCTCGTCGACGACACTGTTTCTTCTCGTTCGCTCGGGTTT	3033
Qy	960	ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal	979
Db	3034	TGGTACGAGGGACACTTCTGTGTCACCAAGTAGATGACATTTTCCGCTCTTGTGTGT	3093
Qy	980	PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp	999
Db	3094	TTTCGCGAGATTCTCTTGTGTCATCCGCGGCACCGCTCTTTCCTTTGACACGAC	3153
Qy	1000	TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu	1019
Db	3154	ATGGCAAGCGGAAGATCGCGCGCGGAATCCGACGACTGTTCCGACGGAAGCCACA	3213
Qy	1020	IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe	1039
Db	3214	ATTGATAACTGGTCTGAAGAGCGCGAAGCTCGAAACGGTGGGAAGTGAATCGAAAT	3273
Qy	1040	AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer	1059
Db	3274	AGGAACGTGCACCTTCAGATACCCGACCCGCCACAGACAGCTGTCTCGCGGGCTTG	3333
Qy	1060	LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer	1079
Db	3334	CTGACCGTGAAGCGTGACAAATATGTGTGCGCTGTGCGACCCAGCGGTGTGGCAAG	3393
Qy	1080	ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp	1099
Db	3394	ACCACCAATGCATTGCTTGAGCGCTTTACGATGCGATTGCGGGTCCATCTTGTTGAT	3453
Qy	1100	GlyLysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSer	1119
Db	3454	GGGAAGGACATAAGTAAACTAATATCAACTCCTACCAGCAGCTTCTGCTCACTGGTC	3513
Qy	1120	GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer	1139
Db	3514	CAGGAGCGCACTGTACCAAGCGCACCATCAAGAAAACATCTTACTTGTGTTGTCGAA	3573
Qy	1140	ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe	1159
Db	3574	CATCAGCTACCGNAGAAATCTTGATTAGGCTTGCAGGAGCGCTAATATCTACGACTC	3633
Qy	1160	IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer	1179
Db	3634	ATCATGTGCTCCGAGGGCTTTAATACAGTTGTTGGCAGCAGGAGGAGCATGTGCT	3693
Qy	1180	GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu	1199
Db	3694	GGCGGCAAAAGCAACGTGTGGCCATGTGCCGAGCCCTTCTTCGGGATCCCAAAATC	3753
Qy	1200	LeuLeuAspGlnAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla	1219
Db	3754	CTTCTCGATGAAGCGAGTTCAGCCCTCGACTCCGAGTCAGAAAGGTCTGTCAGGCG	3813
Qy	1220	LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle	1239
Db	3814	TTGGATGCGCTCCCGAGCGCCGAACCAATCGCGGTTCACACCGACTCAGCACCAT	3873
Qy	1240	GlnAsnAlaAspLeuIleValPheGlnAsnGlnLysValLysGluHisGlyThrHis	1259
Db	3874	CAAAAGCGGACGTTATCTATGTTTCGACCAAGCAGATCGTCAAGAGCGGAACGAC	3933
Qy	1260	GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla	1279
Db	3934	AGCGAGCTGGTCCAGAAAAGCGCGGTACTACGAGCTGGTCAACTCGACAGCTGGCG	3993

QY 1280 Lys 1280
Db 3994 AAG 3996

RESULT 14

```

US-08-996-545-3
; Sequence 3, Application US/08996545
; Patent No. 5928898
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene at rD of
; TITLE OF INVENTION: Aspergillus Nidulans
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,545
; FILING DATE:

```

Alignment Scores:		
Pred. No.:	4,95e-372	4002
Score:	2481.00	538
Percent Similarity:	59.41%	223
Best Local Similarity:	42.00%	458
Query Match:	38.33%	62
DG:	2	8
		Gaps:
		Length:
		Matches:
		Conservative:
		Mismatches:
		Indels:

US-09-672-725C-25 (1-1281) x US-08-996-545-3 (1-4002)

[illegible]

Db 1435 CGAUUCUACAUGCCUGUGCGGUAUGUUUGUGGUAUGGCAUGACAUCAAGGACCUC 1494
Qy 464 AsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAla 483
Db 1495 AUUCUGCGUGGUGCGCAACAGAUUCUUUGUGUAGCCAGAGCCUGUUUUUGGC 1554
Qy 484 ThrThrIleAlaGlnAsnIleArgTyrGly-----ArgGlnAsnValThr 498
Db 1555 ACCAGAUUUUAAGAAUAUUAAGCCAGGUCUACUGCGGCACAAAGUACGAGAAUAUCC 1614
Qy 499 MetAspGlu-----IleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 514
Db 1615 GAGGAUAAAGUCCGGAACAUCAUGAGAGCGCGCAAAAUAGGCAUUGCUCAUGACUUU 1674
Qy 515 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSer 534
Db 1675 AUUACUGCCUGGCGAAGGUUAUGAGACAAUUGUGGCGGCGGUUUUCCUUUCA 1734
Qy 535 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu 554
Db 1735 GGUGCCAGAAACAGCGCAUUGGCAUUGCCGUGCGGUUUAUGAGACCCAAAUUCCUG 1794
Qy 555 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla 574
Db 1795 CUCCUGGAAGAGCUACUUGCGGUUGGACACAAAUUCCGAGGCGUGGUCAAGCAGCU 1854
Qy 575 LeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrVal 594
Db 1855 UUGAGAGGCGACGUGAAGCCGAACUACUUAUUGUAGUCGUCGCGUUUCCUCCAGAU 1914
Qy 595 ArgAsnAlaAspValIleAlaGlyPheAspGlyValIleValGlnLysGlyAsnHis 614
Db 1915 AAAACGGCGCACACAUUGUGGUUGGUCAUUGGCAAAUUGCUGAACAAAGAACUCAC 1974
Qy 615 AspGluLeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGly 634
Db 1975 GAUGAAUUGUGUAGCCGCGGCGGUUAUUGGCAAAUUGUGGAGGCGUACAUCAAU 2034
Qy 635 AsnGluIleGlu-----LeuGluAsnAlaThrGlyGluSerLysSerGluSerAsp 651
Db 2035 GAACAGGAAGAGUGUGAGCGCCGAGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2094
Qy 652 AlaLeuGluMetSerProLysAspSerGlySerSer-----Leu 664
Db 2095 AUUCCCAAAUACAAACUGGCUAAGCGCAUCAUCCGACGCGAAACCCACAAC 2154
Qy 665 IleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAspArgLysLeu 684
Db 2155 AUUAGCCGACGCGGACCCACAGAGUCUGUUGCCAGCGCGAGUUCUUAAGAGACCC --- 2211
Qy 685 GlyThrLysGluAspLeuAsnGluAsnValProProValSerPheThrArgIleLeuLys 704
Db 2212 -----CCGAAACAAUCCCGAAAUUACUUAUUGGAGCGGCGUCAA 2253
Qy 705 -----LeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db 2254 UUGUGUUCUUCUACACCGCCUGAAAUCCCGUACAUUGCUAUGGCGUUCUUCUCA 2313
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2314 GUGUUAUGUGUGUGGUGGCGCAACCCACGAAAGCGUGCUUAUUGCUAAGCAACGACACA 2373
Qy 741 PheThrArgAspGluAspProGluThrLys---ArgGlnAsnSerAsnMetPheSerVal 759
Db 2374 CUUCGCUCCAGAAUACAAUAUAGCAAGCUUCGACAUAGGCGGAUUUUGGUAUUG 2433
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2434 AUGUUUCUGUGUGGUAUCAUUAUCAGUUUAUCAGCAGUACCAACAAUGGUGUGCAUUU 2493
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2494 GCCGUAGUCCCGAGAGACUUAUUGCGCGGAGAGAGACUGCCUUUUGGAGCAUCUC 2553

Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2554 CGUACAAGACAUUGCUUUGACAAGAGAAUAGCACCGCGCUCUGACGACUUCU 2613
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2614 CUGUCCACCGAGCAAGCAUCUCUGCGGUUAGCGGUGUGACUCUAGGACCAUUCU 2673
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2674 AUGACCUCCAGACCCUAGGAGCGCUUAUCAUUAUUGCCUGGCGAUUGGGAUAUUG 2733
Qy 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2734 GCCUUAUUUAUUGUACUCUGGUGUGCGGUCUCUGGCAUGCGGUUUCUACCGAUUCU 2793
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGlnLeuGluGlyAlaGlyLysIleAla 899
Db 2794 AUGUAGCCCAAGUUUCAAUACAGCUCUCCAAAGCUUGCUUAUGAGGAUUCGCAACUUGCU 2853
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2854 UGCAGGCGCAUCUUGCUUACGACAGUGGCGUUAUACCCGCGGAAGGAUGUGUGG 2913
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2914 GAGAUUUAACGACCCUAGCAGCACAAGGACGACAGUAAUUCUCUGUUGAGG 2973
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2974 UCAUCCCUUGUAUUGGUGCGGACGACUUGUUUCUGGUGGUGGUGGUGGUGGUGG 3033
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 3034 UGUUACGAGGAGGACAUUCUUGGUCACCAAGUAUGACAUUUUCCGUUCUUGUUUGU 3093
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3094 UUCUCCGAGAUUCUUGUGUCUAAUCCGCGGACCGUCUUCUUCUUGCACCAGAC 3153
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3154 AUGGCAAGGCGAAGAAUGCGCGCCGAAUUCGACGACUUGGCGACGAAAGCCACA 3213
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3214 AUUGAUAUCUGUGUAGAGGCGGAGAGCUCGAAACGUGGAGGAGGAGGAGGAGGAG 3273
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3274 AGGAACGUGCACUUCAGAUACCGACCGCCAGAACCGCUGUCGCGGCGGUGGAC 3333
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3334 CUGACCGGAGGAGCCUGGAGCAUAUUGCGGUGGCGGAGGAGGAGGAGGAGGAGGAG 3393
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3394 ACCACCAUUGCAUUGCUUGAGCGUUUUAUGAGUUGCGGUGGUGGUGGUGGUGGUG 3453
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSer 1119
Db 3454 GGGAAGGACAAUAAAUAAUACUACUCCGACCGCUUUCUGUUCUUCUUCUUCUUCU 3513
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3514 CAGGAGCGACACUUGGAGGCGGACCAUAGGAAGAAACAUUUAUUGGUAUUGGUAU 3573
Qy 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
Db 3574 GAUGAGGUACCGGAAGAAUUCUUAUAGGCUUGCAAGGAGCGUAAUUAUCACGACUC 3633

```
QY 1160 ileGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
   ||| ::::::::::: ::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
Db 3634 AUCAUGUCGUCUCCGAGGCUUAUAACAGUUGUUGGCAGAGGAGGCAUGUUGCU 3693

QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3694 GCGGCCCAAAAGCAACGUGGCGCAUUGCCGAGCCUUCUUGGGAUCCCAAAAUCCUU 3753

QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnLuAla 1219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3754 CUUCCUGAUGAAGCAGCGACCCUCCAGCUCGAGUUGUUGGCAAGAAAGUUGCCAGCGCGCU 3813

QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3814 UUGGAUGCGCGUCCCGAGCGGACCAACAAUGCGCGUUGCACACCGACACGACGAUU 3873

QY 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3874 CAAAAGCGGAGGUUAUCUAUGUUUCACCAAGCAAGCAUGUCGAAAGCGGAACGCAC 3933

QY 1260 GlnGlnLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
   ::::::::::: ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3934 ACGAACUGGUCACAAAAGGCGCGUACGAGCGUGGCAACUUGCAGAGCUUGGCG 3993

QY 1280 Lys 1280
   |||
Db 3994 AAG 3996

RESULT 15
US-09-328-320-1
; Sequence 1, Application US/09328320
; Patent No. 6228615
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene atrd of
; ASPERGILLUS NIDULANS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/328.320
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/996,545
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11766
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-2763
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
```

```
;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4002
US-09-328-320-1

Alignment Scores:
Pred. No.: 4 95e-272 Length: 4002
Score: 2481.00 Matches: 538
Percent Similarity: 59.41% Conservative: 223
Best Local Similarity: 42.00% Mismatches: 458
Query Match: 38.33% Indels: 62
DB: 4 Gaps: 8

US-09-672-725c-25 (1-1281) x US-09-328-320-1 (1-4002)
QY 24 LysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPheArgTyrSer 43
   ||| ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 268 AAGACGCGAGCTGGAGGAGATCAAGATAACATCTCTTCTCGTCTCTCGCGGTATGCA 327

QY 44 AsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIleHisGlyAla 63
   ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 328 ACAAAGATGGATATACATTATCATGGTAATCAGTACATCTGTCCATTGCTGCCGCGTCG 387

QY 64 AlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAsnAlaGly 83
   ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 388 ACTTCCAGAGGATAATGTTATAT----- 411

QY 84 IleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsnThrGln 103
   ||| ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 -----CAA 414

QY 104 HisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyrSerGlyIle 123
   ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 415 ATCTCGTACGACGAGTTCTATGATGAATGACCAAGAACGCTACTGCTCTCGTATACCTC 474

QY 124 GlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeuAlaAlaGly 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 475 GGTATCGCGGAGTTGTCTACTGTCTAGTACTGTGGCTTCATCTACCGGAGAA 534

QY 144 ArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGluIleGly 163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 535 CACGCCACGCAAGATCCGCGAGTATTACCTTGAGTCTATCTCGCCACAGACATGGC 594

QY 164 TrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAspValSerLys 183
   ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 TATTTGATAAACTCGGTGCGGGGAAGTACCACCGGTATACACCGCATACAAACCTT 654

QY 184 IleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIleAlaThrPhePhe 203
   ||| ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 ATCCAGATGGCATTCGGAGAGGTCGCTCCTACITTGACTGCCCTGGGCATTCGTG 714

QY 204 ThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIleLeuAlaIle 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 715 ACAGCATTCATTATCGCTACGTCAAATACTGGAAGTTGGCTCTAATTTTCAGCTCAACA 774

QY 224 SerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPheThrAsp 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 775 ATGTGGCCCTCGTCTCACCATGGGCGGTGTTCTCAGTTTATCATCAAGTACACAA 834

QY 244 LysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeuAlaAlaIle 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 835 AAGTCGCTTACAGCTACGTCAGCGCGGCACCTGTTCGGAAGAGGTATCAGCTCCATC 894

QY 264 ArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsnLysAsnLeu 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 895 AGAAATGCCACAGCGTTTGGCACCACCAAGCAAGCTTGCAGAGCAGTATGAGGTCCTTA 954

QY 284 GluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSerIleGlyAla 303
   ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 GACGAGCTGAGAAATGGGAACAAACACAGATTGTCATGGTGTTCATGATTGCGGCC 1014
```

Qy 304 AlaPheLeuLeuLeuTyrAlaSerTyrAlaLeuAlaPheTyrTyrGlyThrSerLeuVal 323
Db 1015 ATGTTTGGCCTTATGTAACGAACTTCTGGCTTCTGGATGGTCTCTGTTCCCTG 1074
Qy 324 LeuSerSerGluTyrSerIleGlyGlnValLeuThrValPhePheSerValLeuIleGly 343
Db 1075 GTAGATGGTGCAGTCGATGTGGTGATATCTCACAGTTCTCATGGCCATCTTGATCGGA 1134
Qy 344 AlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArgGlyAla 363
Db 1135 TCGTTCTCTTGGGAAAGCTTAGTCCAAATGCTCAAGCAATTTACAAACGCTGTGGCCGG 1194
Qy 364 AlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSerLysSer 383
Db 1195 GCCCAAGATATTGGAACGATCGATCGCCAGTCGCCATTTAGATCCATATTGCAACGAA 1254
Qy 384 GlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPheSerTyr 403
Db 1255 GGAAGACGCTCGACCATTTTGGGCGCACATTTAGGTACGCAATGTCAAGCATATTATAC 1314
Qy 404 ProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSerGlyGln 423
Db 1315 CCATCTAGACCCGAGCTACCGTCATGGAGGATGTTCTCTGTCAATGCCCGCTGGAAA 1374
Qy 424 ThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrValGlnLeuMetGln 443
Db 1375 ACAACCGCTTAGTCGGCCCTCTGGCTCTGGGAAAGTACGGTGGCTGGTTGAT 1434
Qy 444 ArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArgThrIle 463
Db 1435 CGATTCTACATGCTGTTCCGGGTACGGTTTGTGGATGGCCATGACATCAAGGACCTC 1494
Qy 464 AsnValArgHisArgGluIleThrGlyValValSerGlnGluProValLeuPheAla 483
Db 1495 AATCTCGCTGGCTTCGCCAACAGATCTCTTGGTTAGCCAGGACCTCTCTTTTGGC 1554
Qy 484 ThrThrIleAlaGluAsnIleArgTyrGly-----ArgGluAsnValThr 498
Db 1555 ACGACGATTTATAAGAATATTAGGCACTTAGCCGCTCATCGGCACAAAGTACGAAATGCC 1614
Qy 499 MetAspGlu-----IleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 514
Db 1615 GAGATAGTCCGGAACTCATCGAGACCGCGCAAAATGCGGAATGCTCATGACTTT 1674
Qy 515 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSer 534
Db 1675 ATTACTGCTTGCCTCAAGTTATGAGACCAATGTTGGCAGCGTGGCTTCTCCTTTCA 1734
Qy 535 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu 554
Db 1735 GGTGGCCAGAAACAGCGCATTTGCAATCGCCGCTGGCGTTGTTAGTCACCAAAATCCTG 1794
Qy 555 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla 574
Db 1795 CTCTCGATGAAGCTACTTCGGCTTGGACACAAATCGAAGGCGTGGTTCAACGAGCT 1854
Qy 575 LeuAspLysAlaArgLysGlyArgThrIleValIleAlaHisArgLeuSerThrVal 594
Db 1855 TTGGAGAGGCGACGCTCAAGCGCAACTATTGTGATCGCTCATCGCTTCCACGATC 1914
Qy 595 ArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHis 614
Db 1915 AAAACGGCCACAACTTGTGTTCTTGGTCAATGGCAAAATTTGCTGAACGAAGAACTCAC 1974
Qy 615 AspGluLeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGly 634
Db 1975 GATGAATTGGTTCACCGCGGAGCGCTTATCGCAAACTTGTGGAGGCTCAACGTATCAAT 2034
Qy 635 AsnGluIleGlu-----LeuGluAsnAlaThrGlyGluSerLysSerGluSerAsp 651
Db 2035 GAACAGAAGGAAGCTGACCGCTTGGAGGACCGCGAGCTGAGGATCTCAGCAATCGAGAT 2094
Qy 652 AlaLeuGluMetSerProLysAspSerGlySer-----Leu 664

Db 2095 ATTGCCAAATCAAAACTCGCTCAAGCGCATCATCCGATCTCGACGGAACCCACAAACC 2154
Qy 665 IleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAspArgLysLeu 684
Db 2155 ATTGACCCAGCGGACCCCAAGTCTGTTCACGCGGATCTTCTTAAAGACCC--- 2211
Qy 685 GlyThrLysGluAspLeuAsnGluAsnValProValSerPheThrArgIleLeuLys 704
Db 2212 -----CCCGAAACAACCTCCGAAATACTCATTTATGACGCTCTCTCAAA 2253
Qy 705 -----LeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db 2254 TTTGTTGCTTCTCTCAACCGCCTGAAATCCGTACATGCTCATCGTCTGTCTTCTCA 2313
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2314 GTGTTAGCTGGTGGTGGCCAAACCCAGCAGTGTCTATATGCTAAAGCCATCAGCACA 2373
Qy 741 PheThrArgAspGluAspProGluThrLys---ArgGlnAsnSerAsnMetPheSerVal 759
Db 2374 CTCTCGCTCCAGAAATCACAATATAGCAAGCTTCGACATGATGGGATTTCTGGTCATTG 2433
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlnPheThrPhe 779
Db 2434 ATGTTCTTCGTGGTGGTATCATTTACGTTATCACGAGTCAACCAATGGTGTGTCATT 2493
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2494 CCGTATGCTCCGAGAGACTTATTCGTCGCCGAGAGACTGCTTCGGACGACTCTC 2553
Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2554 CGTCAAGACATGCTTCTTTGACAAGGAAGAATAGCACCGCGCTCTGACCTCTTC 2613
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2614 CTGTCACCGAGAGCAAGCATCTCTCCGGTGTAGCCGCTGTGACTCTAGGCACGATCTG 2673
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2674 ATGACCTCCAGCCCTAGGAGCGGCTATCATTTATGCCCTGGCGATGGGTGAAATTG 2733
Qy 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2734 GCCTTAGTTTGTATCTCGTGTGCGGTTCTCTCTGTCGCGTCTTCTACCCGATCTAT 2793
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2794 ATGTAGCCCAAGTTCAATCACGCTCCAGCTTCCTATGAGGATCTGCAAACTTTGCT 2853
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2854 TGCAGGCTACATGCTCTATCCGACAGTTCGCTCATTAACCCGGGAAGGATGTCTGG 2913
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2914 GAGATTACCATGCCAGCTTGAGCGACAGAGCAGACCACTTAATCTCTGCTCTGAGG 2973
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2974 TCATCCCTGTTATATGCTGTCGACGACTTGTTCCTTCGCTGCGTGGCTCGGCTT 3033
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 3034 TGGTACGGAGGACACTTCTTGGTCACACGAGTATGACATTTTCCGCTCTCTTTGTTGT 3093
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3094 TTCTCCGAGATCTCTTTGGTGTCAATCGCGGCGACCTCTCTTTCTTTCGACACAGAC 3153
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019

Db 3154 ATGGCAAGGGGAGAGATCGGCCGCCGAATTCGACGACTGTTCGACCGAAAGCCACAA 3213
QY 1020 ileAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3214 ATGTATACTGGTCTGAAGAGGGCGAGAGCTCGAAACGGTGAAGGTGAATCGAATTT 3273
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3274 AGGAACGTGCATTCAGATACCGACCGCCGACAGACAGCCTGTCTCGCGGGCTTGGAC 3333
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3334 CTGACCGTGAAGCCCTGGACAATATGTTGCGCTTGTGGACCCAGCGGTTGTGCAAGAGT 3393
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3394 ACCACCATGTGATGCTTGACGCGCTTTTACGATCGGATGCGGGGTCCATCTTGTGTGAT 3453
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3454 GGAAGGACATAAGTAAATCAACTCACTCTACCGCAGCTTCTGTCACTGCTCAGC 3513
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3514 CAGGAGCGGACACTGTACCGAGGCGCCATCAAGGAAACATCTTACTTGTGTTATGTCGAA 3573
QY 1140 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3574 GATGACGTACCGGAGAAATTCCTTGATTAAAGCTTGCAGGACGCTAATATCTACGACTTC 3633
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3634 ATCATGTGCTCCCGAGGGGCTTAAATACAGTTCTTGGCAGCAAGGAGGCATGTTGTCT 3693
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3694 GCGGCCCAAAAGCAACGTTGCGCATTCGCCGAGCCCTTCTTCGGGATCCCAAAATCCTT 3753
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3754 CTTCTCGATGAAGCGAGCTCAGCCCTCGACTCCGAGTCAGAAAAGTCCGTCACGGCGGCT 3813
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3814 TTGGATGCCGTGCCCGAGGCGGAACCAACATCCCGCTTGCACACCGACTCAGCACGATT 3873
QY 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3874 CAAAAGCGGACGTTATCTATGTTTTCGACCAAGCAAGATCGTCGAAGCGGAACGCAC 3933
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3934 AGCGAAGTGGTCCAGAAAAGGCGGCTACTACGAGTGTGTCACTTGCAGAGCTTGGGC 3993
QY 1280 Lys 1280
Db 3994 AAG 3996

Search completed: November 6, 2002, 19:28:53
Job time : 382.556 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:33:09 ; Search time 20.063 Seconds
(without alignments)
6135.192 Million cell updates/sec

Title: US-09-672-725C-25

Perfect score: 6473

Sequence: 1 MDPEGGRKGSARFNWKGK.....LLAQKGYFSMVSQAGAKR 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5866	90.6	1280	1 DVHU1	multidrug resistan
2	5656.5	87.4	1276	1 DVHV1C	multidrug resistan
3	5638	87.1	1276	2 A34786	multidrug resistan
4	5301	81.9	1276	1 DVMS1	multidrug resistan
5	5231.5	80.8	1277	2 JH0502	p-glycoprotein - r
6	4995	77.2	1104	1 DVMS1A	multidrug resistan
7	4915.5	75.9	1279	1 DVHU3	multidrug resistan
8	4863	75.1	1276	1 DVMS2	multidrug resistan
9	4822.5	74.5	1281	2 I48123	multidrug resistan
10	4809	74.3	1278	2 S41646	p-glycoprotein iso
11	4432	68.5	1287	2 S55692	p-glycoprotein - r
12	3215.5	49.7	1321	2 T42228	multidrug resistan
13	3211.5	49.6	1321	2 T42842	p-glycoprotein sis
14	2867	44.3	1294	2 T19982	bile salt transpor
15	2822.5	43.6	1275	2 T31073	hypothetical prote
16	2822	43.6	1289	2 D87789	multidrug resistan
17	2744	42.4	1321	2 T23476	protein C34G6.4 li
18	2742	42.4	1321	2 S27337	hypothetical prote
19	2674.5	41.3	655	1 DVHV2C	multidrug resistan
20	2646	40.9	1283	2 A47377	multidrug resistan
21	2605.5	40.3	1286	2 T02187	probable ABC trans
22	2598.5	40.1	1292	2 T48007	P-glycoprotein hom
23	2595.5	40.1	1278	2 E86155	probable ABC trans
24	2562.5	39.6	1302	2 A41249	multidrug resistan
25	2526	39.0	1229	2 D85023	p-glycoprotein-lik
26	2523	39.0	1229	2 T52319	P-glycoprotein-lik
27	2521	38.9	1230	2 E85023	probable P-glycopr
28	2493	38.5	1302	2 B41249	multidrug resistan
29	2478	38.3	1229	2 F86155	probable ABC trans

30	2446	37.8	1408	2 T43261	multidrug resistan
31	2436.5	37.6	1323	2 H85202	hypothetical prote
32	2424	37.4	1286	2 A42150	p-glycoprotein pp
33	2422	37.4	1310	2 S30328	multidrug resistan
34	2412	37.3	1266	2 T22090	hypothetical prote
35	2394.5	37.0	1268	2 T22094	hypothetical prote
36	2387	36.9	1302	2 S30327	multidrug resistan
37	2341.5	36.2	1254	2 S27338	p-glycoprotein C -
38	2329.5	36.0	1222	2 T14805	hypothetical prote
39	2304	35.6	1233	2 T4251	p-glycoprotein 2 -
40	2297	35.5	1245	2 G86404	probable P-glycopr
41	2292	35.4	1254	2 T30855	multidrug resistan
42	2288	35.3	1307	2 T30882	multidrug resistan
43	2275	35.1	1362	2 T41534	leptomycin B resis
44	2270.5	35.1	1318	2 T21266	hypothetical prote
45	2248	34.7	1347	2 F86405	probable P-glycopr

ALIGNMENTS

RESULT 1

DVHU1

multidrug resistance protein 1 - human

N:Alternate names: P-glycoprotein 1

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1990 #sequence_revision 18-Aug-1995 #text_change 19-Jan-2001

C:Accession: A34914; PS0162; S15500; A25059; S43838; I52238; I65204

R:Chen, C.; Clark, D.; Ueda, K.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.

J. Biol. Chem. 265, 506-514, 1990

A:Title: Genomic organization of the human multidrug resistance (MDR1) gene and origin of the gene

A:Reference number: A34914; MUID:90094448

A:Accession: A34914

A:Molecule type: DNA

A:Residues: 1-1280 <CHE>

A:Cross-references: GB:M29447; GB:J05168; NID:gl87496; PIDN:AAA59576.1; PID:g386862

R:Kioka, N.; Yamano, Y.; Komano, T.; Ueda, K.

submitted to JIPID, April 1991

A:Reference number: PS0162

A:Accession: PS0162

A:Molecule type: DNA

A:Residues: 1-22 <KIO>

R:Kioka, N.; Yamano, Y.; Komano, T.; Ueda, K.

submitted to the EMBL Data Library, April 1991

A:Description: Transcriptional regulation of multidrug resistance gene (MDR1) expressed

A:Reference number: S15500

A:Accession: S15500

A:Molecule type: DNA

A:Residues: 1-22, 'R' <K12>

A:Cross-references: EMBL:X58723; NID:g34522; PIDN:CAA41558.1; PID:g34523

R:Chen, C.; Chin, J.E.; Ueda, K.; Clark, D.P.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.

Cell 47, 381-389, 1986

A:Title: Internal duplication and homology with bacterial transport proteins in the m

A:Reference number: A25059; MUID:87028230

A:Accession: A25059

A:Molecule type: mRNA

A:Residues: 1-184, 'V', 186-1280 <CH2>

A:Cross-references: GB:M14758; NID:gl87468; PIDN:AAA59575.1; PID:g307180

R:Chambers, T.C.; Pohl, J.; Glass, D.B.; Kuo, J.F.

Biochem. J. 299, 309-315, 1994

A:Title: Phosphorylation by protein kinase C and cyclic AMP-dependent protein kinase

A:Reference number: S43838; MUID:94220047

A:Accession: S43838

A:Molecule type: protein

A:Residues: 656-689 <CHA>

R:Geckeler, V.; Weger, S.; Probst, H.

Biochem. Biophys. Res. Commun. 169, 796-802, 1990

A:Title: mdrl/p-glycoprotein gene segments analyzed from various human leukemic cell

A:Reference number: I52238; MUID:90290529

A:Accession: I52238

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 178-215 <RES>

A:Cross-references: GB:M37724; NID:g183537; PIDN:AAA8047.1; PID:g553314
A:Accession: I65204
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 800-856 <RE2>
A:Cross-references: GB:M37725; NID:g183538; PIDN:AAA8048.1; PID:g553315
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: GDB:PGY1; MDRI
A:Cross-references: GDB:120712; OMIM:171050
A:Map position: 7q21-7q21
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; p-loop; phosphoprotein;
F:1-638/Domain: hydrophobic <HBL>
F:49-350/Domain: hydrophobic <HBL>
F:351-637/Domain: hydrophobic <HBL>
F:410-604/Domain: ATP-binding cassette homology <ABC1>
F:427-434/Region: nucleotide-binding motif A (p-loop)
F:551-555/Region: nucleotide-binding motif B
F:638-708/Domain: linker <LIN>
F:709-993/Domain: hydrophobic <HB2>
F:994-1280/Domain: hydrophobic <HL2>
F:1053-1249/Domain: ATP-binding cassette homology <ABC2>
F:1070-1077/Region: nucleotide-binding motif A (p-loop)
F:1196-1200/Region: nucleotide-binding motif B
F:91,94,99/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:433/Binding site: ATP (Lys) #status predicted
F:661,667,671/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status exp
F:667,671,683/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status
F:1076/Binding site: ATP (Lys) #status predicted

Query Match 90.6%; Score 5866; DB 1; Length 1280;
Best Local Similarity 90.8%; Pred No. 0;
Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;

QY 1 MDPEGGRKGS-A-EKNFWKMGKSKKKEKKPTVSTFAMFRYSNNWLDRLYMLVGTMAAI 59
DB 1 MDLEGRNGAKKNEFKLNKS-EKDKKEKPTSVFSNFRYSNNWLDRLYMLVGTMAAI 59
QY 60 IHGAALPLMLVPGNMTDSFANNKTSRNKTFPVINESITNTQHTFINHLEEMTYAY 119
DB 60 IHGAGLPLMLVPGEMTDIFANAG-NLEDLMSNITNRSINDNTGPFMN-LEEDMTRIAY 117
QY 120 YSGIGAGLVAAIYQVSWFCLAAAGRIQKIRKOFFHAIIMRQEIWGFVDHVGELNRLFD 179
DB 118 YSGIGAGLVAAIYQVSWFCLAAAGRIQKIRKOFFHAIIMRQEIWGFVDHVGELNRLFD 177
QY 180 DVSKINEGIDKIGMFTQSTATFTTGFIVGFTRGWKLTLVLAISPVLGSLAAIWAKILS 239
DB 178 DVSKINEGIDKIGMFTQSMATFTTGFIVGFTRGWKLTLVLAISPVLGSLAAIWAKILS 237
QY 240 SFTDKELLAYAKAGAAVEELAIRVIAFGGOKKELERNKNLEAKGIGIKKAITANI 299
DB 238 SFTDKELLAYAKAGAAVEELAIRVIAFGGOKKELERNKNLEAKGIGIKKAITANI 297
QY 300 SIGAAFLIIYASALAFWYGTSLVSESYSGQVLTVFFSVLIGAFSGQASPSIEAFAN 359
DB 298 SIGAAFLIIYASALAFWYGTSLVSESYSGQVLTVFFSVLIGAFSGQASPSIEAFAN 357
QY 360 ARGAAEYIFKIDNKPSIDSYSGKHGPNKIDNLEFKNVHFSYPSRKEVKILKGLNLKY 419
DB 358 ARGAAEYIFKIDNKPSIDSYSGKHGPNKIDNLEFKNVHFSYPSRKEVKILKGLNLKY 417
QY 420 QSGGTVALVNSGCGSTTVQLMORLYDPTDGMVCIDGODIRINVRHLREITGVVSQEP 479
DB 418 QSGGTVALVNSGCGSTTVQLMORLYDPTDGMVSDGQDIRINVRHLREITGVVSQEP 477
QY 480 VLFATTIAENIRYGRNVMTDEIAKVEKANYDFIMKLPNKFDTLVGERGAOLSGGQK 539
DB 478 VLFATTIAENIRYGRNVMTDEIAKVEKANYDFIMKLPNKFDTLVGERGAOLSGGQK 537
QY 540 RIATARALVRNPKILLDDEATSDLTSEAVQVVALDKARKGRTTIVIAHRLSTVRNADY 599

DB 538 RIATARALVRNPKILLDDEATSDLTSEAVQVVALDKARKGRTTIVIAHRLSTVRNADY 597
QY 600 IAGFDGVIYVEKGNHDELMKEGIYFKLVTMQTRGNEIELENATGESKSEDALEMSPKD 659
DB 598 IAGFDGVIYVEKGNHDELMKEGIYFKLVTMQTRAGNEVELENADESKSEIDALESSND 657
QY 660 SGSSLIKRSTRRSIHAPOQDQDRKLGTKEBNDLNVPPVSWRILKLNSTWPFVVGIFC 719
DB 658 SRSLIRKRSTRRSVRGSOAQDRKLSTKEALDESIPVSWRIMKLNLTWPFVVGIFC 717
QY 720 AIINGGLQAPAFSIIFRIIGIFTRDEDPETKRQNSNMFSLFLVLGLIISITFEFLOGFTF 779
DB 718 AIINGGLQAPAFSIIFRIIGIFTRIDDPETKRQNSNLSLFLALGIIISITFEFLOGFTF 777
QY 780 GKAGEILTKLRYMVFRSMLRQDVSFEDDPKNTTGALTTRLANDAAQVKAIGASRLAVIT 839
DB 778 GKAGEILTKLRYMVFRSMLRQDVSFEDDPKNTTGALTTRLANDAAQVKAIGASRLAVIT 837
QY 840 QNIANLGTGIIISLIYQWQLTLLALLAIVPIIATAGVVMKMLSGOALKDKKELEGAGKIA 899
DB 838 QNIANLGTGIIISLIYQWQLTLLALLAIVPIIATAGVVMKMLSGOALKDKKELEGAGKIA 897
QY 900 TEAIENFTVVSILTREOKFEYMYAQSLOQVYRNSLRKAHIFGVSFSTQAMWYFSYAGCF 959
DB 898 TEAIENFTVVSILTREOKFEYMYAQSLOQVYRNSLRKAHIFGIIFSTQAMWYFSYAGCF 957
QY 960 RFGAYLVANEPNMFQDVLIVFSAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPL 1019
DB 958 RFGAYLVAHKLMSEFDVLIVFSAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKTPL 1017
QY 1020 IDYSYSPHGLKPTLEGNVTFNEVFNTPRDPIDPVLQGLSLEVKKGTALVVGSSGCGKS 1079
DB 1018 IDYSYSTEGLMPTLEGNVTFNEVFNTPRDPIDPVLQGLSLEVKKGTALVVGSSGCGKS 1077
QY 1080 TVQOLLERFYDPLAGSLVLDGKEIKHLNVQWLRHAHLGTVSOEPIPLFDCSTAEINAYGNS 1139
DB 1078 TVQOLLERFYDPLAGSLVLDGKEIKHLNVQWLRHAHLGTVSOEPIPLFDCSTAEINAYGNS 1137
QY 1140 RVVSHETQAAKEANIHHFTETLPEKYNTRVGDGKQTLQSGGQKORIAIARALVRQPHIL 1199
DB 1138 RVVSOEIVRAAKEANIHAFTESLPNKYSTKVGDKQTLQSGGQKORIAIARALVRQPHIL 1197
QY 1200 LDEATSDLTSEKVVQEQALDKAREGRTCIIVIAHRLSTIONADLIIVVFONGKVEHGTH 1259
DB 1198 LDEATSDLTSEKVVQEQALDKAREGRTCIIVIAHRLSTIONADLIIVVFONGKVEHGTH 1257
QY 1260 QOLLAOKGIYFSWVSVOAGAKR 1281
DB 1258 QOLLAOKGIYFSWVSVOAGTKR 1279

RESULT 2
DVHYIC
multidrug resistance protein 1 - Chinese hamster
N:Alternate names: P-glycoprotein pgpi
C:Species: Cricetus griseus (Chinese hamster)
C:Date: 31-Dec-1990 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A38696; C38696; A27126; S33768; I52823
R:Devine, S.E.; Hussain, A.; Davide, J.P.; Melera, P.W.
J. Biol. Chem. 266, 4545-4555, 1991
A:Title: Full length and alternatively spliced pgp-1 transcripts in multidrug-resista
A:Reference number: A38696; MUID:91154265
A:Accession: A38696
A:Molecule type: mRNA
A:Residues: 1-1276 <DB>
A:Cross-references: GB:M59253; NID:g191154; PIDN:AAA37004.1; PID:g191155
A:Accession: C38696
A:Molecule type: mRNA
A:Residues: 108-1276 <DEF>
A:Cross-references: GB:M59254; NID:g191156; PIDN:AAA37005.1; PID:g191157
A:Experimental source: clone ADX185
A:Accession: B38696

A:Molecule type: mRNA
A:Residues: 1-32,771-1276 <DE2>
A:Cross-references: GB:M59252; NID:g191152; PIDN:AAA37003.1; PID:g191153
A:Experimental source: Clone ADX124
R:Endicott, J.A.; Juranka, P.F.; Sarangi, F.; Gerlach, J.H.; Deuchars, K.L.; Ling, V.
Mol. Cell. Biol. 7, 4075-4081, 1987
A:Title: Simultaneous expression of two P-glycoprotein genes in drug-sensitive Chinese H
A:Reference number: A27126; MUID:88122132
A:Accession: A27126
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 706-1276 <END>
A:Cross-references: GB:M17897; NID:g191158; PIDN:AAA37006.1; PID:g191159
R:Zastawny, R.L.; Ling, V.
Biochim. Biophys. Acta 1173, 303-313, 1993
A:Title: Structural and functional analysis of 5' flanking and intron 1 sequences of the
A:Reference number: S33768; MUID:93305724
A:Accession: S33768
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-21 <ZAG>
A:Cross-references: EMBL:L03286
R:Teeter, L.D.; Eckersberg, T.; Tsai, Y.; Kuo, M.T.
Cell Growth Differ. 2, 429-437, 1991
A:Title: Analysis of the Chinese hamster P-glycoprotein/multidrug resistance gene pgp1
A:Reference number: I52823; MUID:92088970
A:Accession: I52823
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-21 <RES>
A:Cross-references: GB:S81975; NID:g240862
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic anticancer drugs.
C:Genetics:
A:Gene: pgp1
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: alternative splicing; ATP; duplication; glycoprotein; nucleotide binding; P-
F:407-601/Domain: ATP-binding cassette homology <ABC1>
F:424-431/Region: nucleotide-binding motif A (P-loop)
F:548-552/Region: nucleotide-binding motif B
F:1050-1246/Domain: ATP-binding cassette homology <ABC2>
F:1067-1074/Region: nucleotide-binding motif A (P-loop)
F:1193-1197/Region: nucleotide-binding motif B
F:87,91,96/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:430/Binding site: ATP (Lys) #status predicted
F:1073/Binding site: ATP (Lys) #status predicted
Query Match 87.4%; Score 5656.5; DB 1; Length 1276;
Best Local Similarity 86.9%; Pred. No. 0;
Matches 1114; Conservative 77; Mismatches 84; Indels 7; Gaps 3;
QY 1 MDPEGRGRKSAKFNWKGKKKKKKKPTVTSTFAMPRYSNWLDRMLVGYTMAAI 60
Db 1 MEFEDESGRKDNFLKMGKRS-KKEKKEKPVVSTFMPRYAGWLDRLMLVGYTMAII 59
QY 61 HGAALPLMLVFGNMTDSFANAG-ISRNTKTPVVIINESITNTQHFTHLEEMTYAY 119
Db 60 HGVALPLMLVFGDMTDSFASVGNIPNTAT-----NNATQVNASDFGKLEEMTYAY 114
QY 120 YSGTGAGVLVAAVYQVSWFCLAGRQILKTRKQFFHAIMQEIGWFDVHDVGEINRLTD 179
Db 115 YTGIGAGVLVAVYQVSWFCLAGRQILKTRKQFFHAIMQEIGWFDVHDVGEINRLTD 174
QY 180 DVSKINGIGDKGMFFQSTATPTGTFVGTFRGKLTILVLAISPVGLSAAIWAAIL 239
Db 175 DVSKINGIGDKGMFFQATATPTGTFVGTFRGKLTILVLAISPVGLSAAIWAAIL 234
QY 240 SFTDKELAYAKAGAAVEVLAARTVIAFGQKKELRYNNKLEAKGIGIKKATANI 299
Db 235 SFTDKELQAYAKAGAAVEVLAARTVIAFGQKKELRYNNKLEAKRIGIKKATANI 294
QY 300 SIGAFLLIYASVALAFWYGTSLVLSSEYSIGQVLTAVFFSVLIGAFSIGQASPSIEAFAN 359
Db 300 SIGAFLLIYASVALAFWYGTSLVLSSEYSIGQVLTAVFFSVLIGAFSIGQASPSIEAFAN 359

Db 295 SMGAFLLIYASVALAFWYGTSLVLSSEYSIGQVLTAVFFFAVLAPFSIGQASPSIEAFAN 354
QY 360 ARGAAVEIFKIIDKNPKSIDSYSKSGHKPNKGNLEFKNVHFSPYSPKSKYKILKGLNLKV 419
Db 355 ARGAAVEIFNIIIDKNPKSIDSFSGKNGYKPNKGNLEFKNVHFSPYSPKSKYKILKGLNLKV 414
QY 420 QSGQTVALVGNSSCGGKSTTVQLMORLYDPPDGMVCIIDGQDIRTNVHRLREITGVVQSEP 479
Db 415 QSGQTVALVGNSSCGGKSTTVQLLQRLYDPTGEGVVSIDGQDIRTNVHRLREITGVVQSEP 474
QY 480 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKEDTTLVGERGALSGGQK 539
Db 475 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKEDTTLVGERGALSGGQK 534
QY 540 RIAARALVRNPKILLDEATSDLTSEAVQVVALDKARKGRTTIVIAHRLSTVNRNADY 599
Db 535 RIAARALVRNPKILLDEATSDLTSEAVQVVALDKARKGRTTIVIAHRLSTVNRNADY 594
QY 600 IAGFDGCVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKD 659
Db 595 IAGFDGCVIVEQGNHHELMREKGIYFKLVMTQTRAGNEIELNGNEVGESKNEIDNLDMSK 654
QY 660 SGSSLIKRSTRRSIHAPOQDQDKLGTEDLNENVPVSWFRILKILNSTEWYFVVGIFC 719
Db 655 SASSLIKRSTRRSIRPHDQDKLSTKEALDEDVPPISFWRILKILNSTEWYFVVGIFC 714
QY 720 AIINGGLOPAFSIIFSRIGITRDEDETPETKRONSNFVLFVLVGLIISFIITFFLOGFTF 779
Db 715 AIYNGALOPAFSIFSVGVFTTRDDETKRHSNLSLLELILGLVIFITFFLOGFTF 774
QY 780 GKAGEILTTLRLYRVFMSMLRODVDFDPKNTGTALTTLANDAAOVKATGTSRLAVIT 839
Db 775 GKAGEILTTLRLYRVFMSMLRODVDFDPKNTGTALTTLANDAAOVKATGTSRLAVIT 834
QY 840 QNTANLGTGIIISLIYGWLTLLLAIVPIIAIAGVYEMKLSGQALKDKKELEGAKTIA 899
Db 835 QNTANLGTGIIISLIYGWLTLLLAIVPIIAIAGVYEMKLSGQALKDKKELEGAKTIA 894
QY 900 TEAIENFRVWVSLTREQKEFYAYQSLQVYPYNSLRKAHIFGVSPSITQAMMYFSYAGCF 959
Db 895 TEAIENFRVWVSLTREQKEFYAYQSLQVYPYNSLRKAHIFGVSPSITQAMMYFSYAGCF 954
QY 960 RFGAYLVANEFMFDVLFVSAIVFGAMAVGVSSFPADYAKAKVSAHVIWITEKSPL 1019
Db 955 RFGAYLVARELMTFENVLLVFSALVFGAMAVGVSSFPADYAKAKVSAHVIWITEKSPL 1014
QY 1020 IDSYPHGLKPNTEGNTVFNEVFNYPTRDIPVLQGLSLEVKKGQTLALVSSGCGKS 1079
Db 1015 IDSYSTGLKPNTEGNTVFNEVFNYPTRDIPVLQGLSLEVKKGQTLALVSSGCGKS 1074
QY 1080 TVVOLLERFVDPPLAGSVLDGKEIKHLNVOHLRAHLGIVSQEPILFDCSIAENIAYGDS 1139
Db 1075 TVVOLLERFVDPPLAGSVLDGKEIKHLNVOHLRAHLGIVSQEPILFDCSIAENIAYGDS 1134
QY 1140 RVVSHEIEQAARKEANIHFIETLPEKYNTRYVGDGKQTLSSGQKQRIARALVVRQPHIL 1199
Db 1135 RVVSQDEIERAARKEANIHFIETLPEKYNTRYVGDGKQTLSSGQKQRIARALVVRQPHIL 1194
QY 1200 LIDEATSDLTSEKVVQVQALDKAREGRTCTIVIAHRLSTIQNALDIVVFNQKVKHEGTH 1259
Db 1195 LIDEATSDLTSEKVVQVQALDKAREGRTCTIVIAHRLSTIQNALDIVVFNQKVKHEGTH 1254
QY 1260 QQLAOKGIYFSMVSVOAGAKR 1281
Db 1255 QQLAOKGIYFSMVSVOAGAKR 1276

RESULT 3

A34786
multidrug resistance protein la - mouse

C:Species: Mus musculus (house mouse)

C:Date: 13-Jul-1990 #sequence_revision 13-Jul-1990 #text_change 02-Feb-2001

C:Accession: A34786; A35671

R;Devault, A.; Gros, P.
Mol. Cell. Biol. 10, 1652-1663, 1990
A:Title: Two members of the mouse mdg gene family confer multidrug resistance with over
A:Reference number: A34786; MUID:90205845
A:Accession: A34786
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1276 <DEV>
A:Cross-references: GB:M30697; NID:g199111; PIDN:AAA39517.1; PID:g387429
R;Hsu, S.I.H.; Cohen, D.; Kirschner, L.S.; Lothstein, L.; Hartstein, M.; Horwitz, S.B.
Mol. Cell. Biol. 10, 3596-3606, 1990
A:Title: Structural analysis of the mouse mdrla (P-glycoprotein) promoter reveals the ba
A:Reference number: A35671; MUID:90287150
A:Accession: A35671
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-525, 'HV', 528-938, 'S', 940-1035, 'V', 1037-1276 <HSD>
A:Cross-references: GB:M33581; NID:g199104; PIDN:AAA39514.1; PID:g387427
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; membrane protein; nucleotide binding; P-loop
F:406-600/Domain: ATP-binding cassette homology <ABCi>
F:423-431/Region: nucleotide-binding motif A (P-loop)
F:547-551/Region: nucleotide-binding motif B
F:1049-1245/Domain: ATP-binding cassette homology <ABC2>
F:1066-1074/Region: nucleotide-binding motif A (P-loop)
F:1192-1196/Region: nucleotide-binding motif B
F:429/Binding site: ATP (Lys) #status predicted
F:1072/Binding site: ATP (Lys) #status predicted

	Query Match	87.1%;	Score 5638;	DB 2;	Length 1276;
	Best Local Similarity	87.1%;	Pred. No. 0;		
	Matches 1116;	Conservative	74;	Mismatches	84;
				Indels	8;
				Gaps	4;
Qy	1	MDPEGKRGSAEKNFMKMGKKKKKKKKKPTVSTFAMFYSNWLDRLYMLVGTMAAII	60		
Dd	1	MELEEDLKGRADKNFSKMGKKS-KKEKKKPAVSVLTMFYGAGWLDRLYMLVGTAAII	59		
Qy	61	HGAALPLMLVFGNWTOSFANAG-ISRNKTFPPVIINESITNNTQHFNIHLEEMTTVAYY	119		
Dd	60	HGVALPLMLIFGDMTDSFASVGNVSKNST-----NMSEADKRAMFAK-LEEEMTTVAYY	113		
Qy	120	YSGTIGAGVLVAAYIQVSFWCLAACRQILTKRQFFHAIMRQEIIGWFDVHDVDELNRLTD	179		
Dd	114	YTGIGAGVLIVAYIQVSFWCLAACRQILTKRQFFHAIMRQEIIGWFDVHDVDELNRLTD	173		
Qy	180	DVSKINEGIDKIGMFFQSIATFFTGFTVGTGRWKLTVLTAISPVLGISAATWAKILS	239		
Dd	174	DVSKINEGIDKIGMFFQAMATFFGGFTIGTRGWKLTVLTAISPVLGISAATWAKILS	233		
Qy	240	SFTDKELLAYAKAGAAVEVLAAITRTVAFGGQKKELERYNNKLEEAKGIGIKKAITANI	299		
Dd	234	SFTDKELHAYAKAGAAVEVLAAITRTVAFGGQKKELERYNNLEEAARLGRIGIKKAITANI	293		
Qy	300	SIGAFAFLIYASYALAFWYGTSLVLSSEYSIGQVLTVPFVSVLIGAFSIGQASPIEAFAN	359		
Dd	294	SMGAFAFLIYASYALAFWYGTSLVLSKEYSIGQVLTVPFVSVLIGAFSVGQASPIEAFAN	353		
Qy	360	ARGAAEYFIKIIDNKPSIDSYSKSGHKPDNTKGNLEKFNHVSYPSPSRKEVKILKGLNKV	419		
Dd	354	ARGAAEYEVKIIDNKPSIDSFSGKHGPDNTIOGNLEKFNHVSYPSPSRKEVQILKGLNKV	413		
Qy	420	QSGQTVALVNGSGCGKSTTVQLMORLYDPTDGMWCIDGQDTRTNVRHLREITGVVSEQP	479		
Dd	414	KSGQTVALVNGSGCGKSTTVQLMORLYDPLDGMYSIDGQDTRTNVRYLREIIGVWSEQP	473		
Qy	480	VLFATTAEINTRYGRENVTWDEIEKAKEANAYDFINKLPNKFDTLVGERGAQLSGGQKQ	539		
Dd	474	VLFATTAEINTRYGREDVTWDEIEKAKEANAYDFIMKLPHQFDTLVGERGAQLSGGQKQ	533		
Qy	540	RIATARALVRNPKILLDDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADY	599		
Dd	534	RIATARALVRNPKILLDDEATSAIDTSEAVVQVALDKAREGRTTIVIAHRLSTVRNADY	593		

A:Title: Cell-specific activity of cis-acting regulatory elements in the promoter of the
A:Reference number: 157510; MUID:91042535
A:Accession: 157510
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-21 <RES>
A:Cross-references: GB:M60348; NID:g199102; PIDN:AAA39513.1; PID:g554199
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic antitumor drugs.

C:Genetics:
A:Gene: mdr1 [pgp1]
A:Introns: 21/2; 38/3; 92/1; 111/2; 175/2; 233/3; 274/2; 332/3; 370/3; 407/3; 449/3; 517
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane pr
F:1-637,653-1276/Region: duplication
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:550-554/Region: nucleotide-binding motif B
F:1051-1247/Domain: ATP-binding cassette homology <ABC2>
F:1068-1075/Region: nucleotide-binding motif A (P-loop)
F:1194-1198/Region: nucleotide-binding motif B
F:73,91,96,103/Binding site: carboxylate (Asn) (covalent) #status predicted
F:432/Binding site: ATP (Lys) #status predicted
F:1074/Binding site: ATP (Lys) #status predicted

Query Match 81.9%; Score 5301; DB 1; Length 1276;
Best Local Similarity 80.3%; Pred. No. 1.6e-299;
Matches 1030; Conservative 130; Mismatches 114; Indels 8; Gaps 5;

QY 1 MDPGGKSGAEKFWKMGKSKKKEKKEKPTVSTFAMFRYSNMLDRLVNLVGTMAAII 60
DB 1 MEFEENLKGRADKNFSGKMGKS-KKEKKEKPAVGFPMFRYADWLKCLMILGTLAAII 59
QY 61 HGAALPLMLVFGNMTDSFANAGTSRNKTPVLIINESITNTQHFH-HLEEEETVYAY 119
DB 60 HGTLPLMLVFGNMTDSFTKAS---ILPSITNGSPNLSLIINSLEEEMAIYAY 116
QY 120 YSGIGAGVLVAAYIQVSWFCLAGRQILKIRKQFFHAIMRQEIGWDFVDVHDELNRLTD 179
DB 117 YTGIGAGVLVAYIQVSLWCLAGRQIHKIRKQFFHAIMRQEIGWDFVDVHDELNRLTD 176
QY 180 DVSKINGIGDKGMFQSTATFTGTFVGTGKWLTVLILASPVLGLSAAIAWAKILS 239
DB 177 DVSKINDIGDKGMFQSTITFLAGFIIGFISGWLTLVLVAPSLIGLSALWAKVLT 236
QY 240 SFTDKELLAYAKAGAAVEVLAAIRTVIAGGOKKELERYKNLEAKGIGIKAITANI 299
DB 237 SFTNKELOAKAGAAVEVLAAIRTVIAGGOKKELERYKNLEAKNNGVGIKAITASI 296
QY 300 SIGAAFLIYASALAFWYGTSLVLSSEYSIGQVLTVFFSVLIGAFSIGQASPSIPAFAN 359
DB 297 SIGIAYLLVYASALAFWYGTSLVLSSEYSIGEVLTVFFSVLIGAFSIGHLAPNIEAFAN 356
QY 360 ARGAAEFIKIIDNKPSIDYSKSGHKPDNKNLEKPNVHFYSPKSKVKILKGLNLKV 419
DB 357 ARGAAEFIKIIDNEPSIDFSFTKGYKPDSDIMGNLEFKNVHFYSPKSEVQILKGLNLKV 416
QY 420 QSGOTVALVNSGCGKSTTVOLMORLYDPTDGMVCIDGDIRTINVRHLREITGVVSOEP 479
DB 417 KSGOTVALVNSGCGKSTTVOLMORLYDPLEGVVSDIGQDIRTINVRHLREITGVVSOEP 476
QY 480 VLFATTIENIRYGRENVMTDEIEKAVKEANAYDFIMKLPNKPDTLVGERGAOLSGGQKQ 539
DB 477 VLFATTIENIRYGRENVMTDEIEKAVKEANAYDFIMKLPNKPDTLVGERGAOLSGGQKQ 536
QY 540 RIAIARLVNPKILLDEATSDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADY 599
DB 537 RIAIARLVNPKILLDEATSDTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADY 596
QY 600 IAGFDGCVIYEGKHNHDELMKEGKGYKFLVMTQTRGNEIELENATGESKESDALEMSPKD 659
DB 597 IAGFDGCVIYEGKHNHDELMKEGKGYKFLVMTQTRGNEIELENATGESKESDALEMSPKD 656

QY 660 SSSGLIKRRSTRRSIIHAPOQDKRLGTGKEDLNENVPVFWRLKLNSTENPVPVVGIFC 719
DB 657 SKSPLI--RSIIYRSVHKQDQERRLSMKAEVDEDVPLVFWRLNLSNENPVLVGVJVC 715
QY 720 AIINGLOPAFSIIFRSIIGIFTRDEDPETKRONSNFVLEVLVLGIIISFIIFFLQGFTF 779
DB 716 AVINGCQIPVFAIVFSRIGVFSRDKDDHETKRONCNLSLSFLVLMGLISFVIVFQGFIF 775
QY 780 GKAGEILTKRLRVFWSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGATGSLAVIT 839
DB 776 GKAGEILTKRVYVWFVKSMLRQDISWDFDHKNSTGSLTTRLASDASVSGAMGARLAVT 835
QY 840 QNTANLGTGIIISLIVGWLTLLALLAIVPIIATAGVEMKMLSGQALKDKKLEAGAKTA 899
DB 836 QNVANLGTGVILSLVIGWLTLLVLIIPVILVGGIIEMLKLSGQALKDKKLEISGKTA 895
QY 900 TEAIENFRTVSLTREQKFEYMYAQSLOVPYRNSLRKAHIFGVFSFISITQAMVFSYAGCF 959
DB 896 TEAIENFRTVSLTREQKFTETMYAQSLOVPYRNMKKAHVFGITFTSQAMVFSYAACF 955
QY 960 RFGAYLVANEFMNFQDVLVLFSAIVFAGAMAVGVSSFAFDYAKAKVSAAHVIMIEKSP 1019
DB 956 RFGAYLVAAQQLMTFENVMVLFSAVFEAGMAAGNTSSFAFDYAKAKVSASHIIRIEKTP 1015
QY 1020 IDSYSPHGLKPNTEGNVTFNEVFNYPTRPDIPVLOGLSLEVKKGOTLALVSSSGCGKS 1079
DB 1016 IDSYSTEGLKPTLLEGNVFNFGVQFNYPTRPNIPVLOGLSLEVKKGOTLALVSSSGCGKS 1075
QY 1080 TVVOLLERYDPLAGSVLIDGKEIKHLNVQWLRHAGLIGVYSQBPILFDCSIAENIAYGDS 1139
DB 1076 TVVOLLERYDPMAGSVFLDGKEIKQLNVQWLRHAGLIGVYSQBPILFDCSIAENIAYGDS 1135
QY 1140 RVYSHEEIMQAKEANIHFIETLPEKYNTVRVGDGKTQLSGGQKQRIATARALVROPHTL 1199
DB 1136 RAVSHEEIVRAAKEANIHQFIDSLPKYNTVRVGDGKTQLSGGQKQRIATARALVROPHTL 1195
QY 1200 LLDEATSAIDTSEKVVQEAALDKAREGRTCIVIAHRLSTIQNALDIVVFPONGKVEHGTH 1259
DB 1196 LLDEATSAIDTSEKVVQEAALDKAREGRTCIVIAHRLSTIQNALDIVVFPONGKVEHGTH 1255
QY 1260 QQLLAQKGIYFSMVSVQAGAKR 1281
DB 1256 QQLLAQKGIYFSM--VQAGAKR 1275

RESULT 5

JH0502

p-glycoprotein - rat

N:Alternate names: multidrug resistance protein mdr1b

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 02-Feb-2001

C:Accession: JH0502; S22353; S22352

R:Silverman, J.A.; Raunio, H.; Gant, T.W.; Thorgeirsson, S.S.

Gene 106, 229-236, 1991

A:Title: Cloning and characterization of a member of the rat multidrug resistance (mdr)

A:Reference number: JH0502; MUID:92039081

A:Accession: JH0502

A:Molecule type: mRNA

A:Residues: 1-1277 <SIL>

A:Cross-references: GB:M62425

R:Deuchars, K.L.; Duthie, M.; Ling, V.

Biochim. Biophys. Acta 1130, 157-165, 1992

A:Title: Identification of distinct P-glycoprotein gene sequences in rat.

A:Reference number: S22351; MUID:92223089

A:Accession: S22353

A:Molecule type: DNA

A:Residues: 1212-1226, 'I', 1228-1277 <DEU>

A:Cross-references: EMBL:X61104; NID:g56890; PIDN:CAAA3416.1; PID:g1334219

A:Accession: S22352

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1212-1226, 'I', 1228-1270, 'SV', 1271-1277 <DE2>

A:Cross-references: EMBL:X61103; NID:g56888; PIDN:CAAA3415.1; PID:g1334218

QY	299	ISIGAAFLLIYASYALAFWYGTSLVLSSEYSIGQVLTFFSVLYICAFSIGOASPIEAF	358
Db	121	ISMGAAFLLIYASYALAFWYGTSLVISKEYSIGQVLTFFSVLYICAFSVGOASPIEAF	180
QY	359	NARGAAYEIFKIIDNKPISIDSYSGHGXPDNIKNLEPKNVHFSYPSRKEYILKGLNLK	418
Db	181	NARGAAYEVFKIIDNKPISIDSFSGHGXPDNIQGNLEPKNIHFSYPSRKEYIQLKGLNLK	240
QY	419	VQSGQTVALVGNSSCGKSTTVQLMQRLYDPTDGMVCDIGQDIRTINVHRLREITGVVVSQE	478
Db	241	VKSGQTVALVGNSSCGKSTTVQLMQRLYDPLDGMVSDIGQDIRTINVRYLREITGVVVSQE	300
QY	479	PVLFATTIAENIRYGRENVTWDETEKAVKEANAYDFIMKLPNKEDTLVGERGAQLSGQK	538
Db	301	PVLFATTIAENIRYGREVDTWDETEKAVKEANAYDFIMKLPNQEDTLVGERGAHVSQK	360
QY	539	QRIATARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNAD	598
Db	361	QRIATARALVRNPKILLDEATSALDTESEAVVQQAALDKAREGRTTIVIAHRLSTVRNAD	420
QY	599	VIAGFDDGVIVYKGNHDELMKEGIYFKLVYTMQTRGNEIELEENATGESKSESDALEMSPK	658
Db	421	VIAGFDDGVIVYQGNHDELMKEGIYFKLVYTMQTAGNEIELGNEACKSKDEIDNLDMSK	480
QY	659	DGSSSLIKRRSTRSIIHAPQODRKLGTCKEDLNENVPVPSFWRILKLNSTWEPYFVGIF	718
Db	481	DSGSSLIKRRSTRKSCGPHDQDRKLSTKEALDEDVPPASFWRILKLNSTWEPYFVGIF	540
QY	719	CAIINGGLQPAFSIIFRIIGIFTRDEDPTKRONSNFSLFVLVGLGIIFTFIFFLQGET	778
Db	541	CAIINGGLQPAFVSIFSKVGVFTNGGPPETQRQNSNLSLFLILGIIIFTFIFFLQGET	600
QY	779	FGKAGEILTKRLTYWTFRSMRLQDVSWFDDPKNTTGALTTRLANDAAGVKGATGSRSLAVI	838
Db	601	FGKAGEILTKRLTYWTFKSMRLQDVSWFDDPKNTGTGALTTRLANDAAGVKGATGSRSLAVI	660
QY	839	TQNTIANLGTGIIISLIYGWOLTLALLAIVPIATAGVVEKMLSGQALKDKKELEGAGKI	898
Db	661	QNTIANLGTGIIISLIYGWOLTLALLAIVPIATAGVVEKMLSGQALKDKKELEGSGKI	720
QY	899	ATEAIENFRVTSLTREQKFEYMYAQSLQVPYRNSLRKAHIFGVFSITQAMMYFSYAGC	958
Db	721	ATEAIENFRVTSLTREQKFEYMYAQSLQIPYRNAMKKAHVFGITFSTQAMMYFSYAC	780
QY	959	FRFGAYLVANEFMFDQVLLVFSIAIVEGAMAVGVQSSFADPYAKAKVSAAHVIMIEKSP	1018
Db	781	FRFGAYLVTOOLMTFENVLLVFSIAIVEGAMAVGVQSSFADPYAKATVSASHIIRIEKTP	840
QY	1019	LIDSYPHGLKPNFTLEGNVTFNVEVFNYPTRPDIPVLQGLSLEVKGQTLALVSSGCGK	1078
Db	841	EIDSYSTQGLKPNMLEGNVQFSGVFNYPTRPSIPVLQGLSLEVKGQTLALVSSGCGK	900
QY	1079	STVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHLGIYSQBPIFLDCSIAENIAYGN	1138
Db	901	STVVOLLERFYDPMAGSVFLDGKEIKQLNVQWLRQAQLGIVSQBPIFLDCSIAENIAYGN	960
QY	1139	SRVVSHEEIMQAAKEANIHFIETLPEKYNTRYVGDGKTQLSGGQKQRIATARALVRQPHI	1198
Db	961	SRVVSHEEIVRAKEANIHQIFDLSPKYNTRYVGDGKTQLSGGQKQRIATARALVRQPHI	1020
QY	1199	LLDDEATSDLTSEKVVQALDKAREGRTCIIVIAHRLSTIQNADLIVVFONGKVKHEGT	1258
Db	1021	LLDDEATSDLTSEKVVQALDKAREGRTCIIVIAHRLSTIQNADLIVVIONGKVKHEGT	1080
QY	1259	HQOLLAQKGIYFSKVSVOAGAKR 1281	
Db	1081	HQOLLAQKGIYFSKVSVOAGAKR 1103	
RESULT 7			
DVRH03			
multidrug resistance protein 3 - human			

N:Alternate names: P-glycoprotein MDR3			
C:Species: Homo sapiens (man)			
C:date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001			
C:Accession: JS0051; S01346; A42213; I38015			
R:van der Bliek, A.M.; Koolman, P.M.; Schneider, C.; Borst, P.			
Gene 71, 401-411, 1988			
A:title: Sequence of mdr3 cDNA encoding a human P-glycoprotein.			
A:Reference number: JS0051; MUID:89138016			
A:Accession: JS0051			
A:Molecule type: mRNA			
A:Residues: 1-1279 <VA1>			
R:van der Bliek, A.M.; Baas, F.; ten Houte de Langer, T.; Koolman, P.M.; van der Velde			
EMBO J. 6, 3325-3331, 1987			
A:title: The human mdr3 gene encodes a novel P-glycoprotein homologue and gives rise			
A:Reference number: S01346; MUID:88111519			
A:Accession: S01346			
A:Molecule type: mRNA			
A:Residues: 856-1093, 'FVDRGFO', 1094-1279 <VA2>			
A:Cross-references: EMBL:X06181; NID:934524; PIDN:CAA29547.1; PID:934525			
R:Lincke, C.R.; Smit, J.J.M.; van der Velde-Koerts, T.; Borst, P.			
J. Biol. Chem. 266, 5303-5310, 1991			
A:title: Structure of the human MDR3 gene and physical mapping of the human MDR locus			
A:Reference number: A42213; MUID:91161629			
A:Accession: A42213			
A:Status: not compared with conceptual translation			
A:Molecule type: DNA			
A:Residues: 1-1279 <LIN>			
R:Smit, J.J.; Mol, C.A.; van Deemter, L.; Wagenaar, E.; Schinkel, A.H.; Borst, P.			
Biochim. Biophys. Acta 1261, 44-56, 1995			
A:title: Characterization of the promoter region of the human MDR3 P-glycoprotein gen			
A:Reference number: I38015; MUID:95200972			
A:Accession: I38015			
A:Status: translated from GB/EMBL/DDBJ			
A:Molecule type: mRNA			
A:Residues: 1-61, 'RGSSRVDLQAC' <RES>			
A:Cross-references: EMBL:X35284; NID:g1006662; PIDN:CA84542.1; PID:g1006663			
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c			
tructurally and functionally unrelated lipophilic antitumor drugs.			
C:Genetics:			
A:Gene: GDB:PGY3; MDR3			
A:Cross-references: GDB:120713; OMIM:171060			
A:Map position: 7q21-7q21			
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology			
F:1-640,653-1279/Region: duplication			
F:412-606/Domain: ATP-binding cassette homology <ABC1>			
F:429-436/Region: nucleotide-binding motif A (P-loop)			
F:553-557/Region: nucleotide-binding motif B			
F:638-694/Domain: linker <LINK>			
F:1052-1248/Domain: ATP-binding cassette homology <ABC2>			
F:1069-1076/Region: nucleotide-binding motif A (P-loop)			
F:1195-1199/Region: nucleotide-binding motif B			
F:91.97/Binding site: carbohydrate (Asn) (covalent) #status predicted			
F:435/Binding site: ATP (Lys) #status predicted			
F:1075/Binding site: ATP (Lys) #status predicted			
Query Match			
Best Local Similarity 75.9%; Score 4915.5; DB 1; Length 1279;			
Matches 975; Conservative 131; Mismatches 163; Indels 19; Gaps 9;			
QY	1	MPEGGRRKG-----SAENFWKMGKSKKKKKKKK-PTVSTFAMFRYSNWLDRLYMLV	53
Db	1	MDLEAKNGTAWRPSTSAEGDF-ELGISQKQRKKTKTVKMIGVLTFRYSQWQDKLFMSL	59
QY	54	GTMAAIIHGAALPLMLVFGNMTDSFANAGISRNKTFPVIINSTNTQHFHLEDEM	113
Db	60	GTIMAIHAGSGLPLMMIVFEGMTDFVDT--AGNFSFPVNSLSLLNPK-----ILEEM	113
QY	114	TTYATYYSGIGAGVLVAAYIQVSWCLAGHQLIKRKQFFHAIMRKQFFHGFVDHVGEL	173
Db	114	TRYATYYSGIGAGVLVAAYIQVSWFTLAAGRIKQIRKQFFHAILRQEIFGFIINDITTEL	173

QY 174 NTRLTDVSKINEIGDKIGMPFQSIATFTTGTFTGTRGKLTILVILAISPVGLSAAI 233
Db 174 NTRLTDVSKINEIGDKIGMPFQSIATFTTGTFTGTRGKLTILVILAISPVGLSAAI 233
QY 234 WAKILSFTDKELLAYAKAGAAVEEVLAAIRTVIAFGGQKKELERYKNLEPAKIGIKK 293
Db 234 WAKILSFTDKELLAYAKAGAAVEEVLAAIRTVIAFGGQKKELERYKNLEPAKIGIKK 293
QY 294 ATANTISGAFLLIYASALAFWGTSLVLSSEVSIGQVLVFFSVLIGAFSIGQASPS 353
Db 294 ATANTISGAFLLIYASALAFWGTSLVLSSEVSIGQVLVFFSVLIGAFSIGQASPS 353
QY 354 IEAFANARGAAYEIFIIDNKPSIDSYSGSHGKPDNIKGNLEFKNVHFSPYRKEVKILK 413
Db 354 IEAFANARGAAYEIFIIDNKPSIDSYSGSHGKPDNIKGNLEFKNVHFSPYRKEVKILK 413
QY 414 GLNLKVSOGTVALVNSGCGSTTVOLMORLYDPPDGMVCIDGQDRTINVRHUREITG 473
Db 414 GLNLKVSOGTVALVNSGCGSTTVOLMORLYDPPDGMVCIDGQDRTINVRHUREITG 473
QY 474 VVSOEPLFATTAEINTRYGRENVTWDETEKAVKEANAYDFIMKLPNKFDTLVGERGAOL 533
Db 474 VVSOEPLFATTAEINTRYGRENVTWDETEKAVKEANAYDFIMKLPNKFDTLVGERGAOL 533
QY 534 SGGQKORIAIARALVRNPKILLDEATSDLTSESEAVVQVQALDKARKGRTTIVIAHRLST 593
Db 534 SGGQKORIAIARALVRNPKILLDEATSDLTSESEAVVQVQALDKARKGRTTIVIAHRLST 593
QY 594 VRNADVTAGDDGVYVEKGNHDLMEKEGIYFKLVMTQTRGNEIELENATGESKESDAL 653
Db 594 VRNADVTAGDDGVYVEKGNHDLMEKEGIYFKLVMTQTRGNEIELENATGESKESDAL 653
QY 654 EMSPKSGSLIKRRSTRRIHAPOGODRKLGTKED-LNENVPVSVFWRILKLNSTEWPY 712
Db 654 EMSPKSGSLIKRRSTRRIHAPOGODRKLGTKED-LNENVPVSVFWRILKLNSTEWPY 712
QY 713 FVVGIFCAIINGLOPAFSIFSRIGIFTRDEDPETKRONSMNFSVLVLGIIISFTTF 772
Db 713 FVVGIFCAIINGLOPAFSIFSRIGIFTRDEDPETKRONSMNFSVLVLGIIISFTTF 772
QY 773 FLOGFTGKAGEILTKRLYVFRSMRLRODVSFDDPKNTGTALTURLANDAQAQVGAIG 832
Db 773 FLOGFTGKAGEILTKRLYVFRSMRLRODVSFDDPKNTGTALTURLANDAQAQVGAIG 832
QY 833 SRLAVITQNIANLGTGIIISLIYGWQLTLLLAIVPIIAIAGVYEMKLSGQALKDKKEL 892
Db 833 SRLAVITQNIANLGTGIIISLIYGWQLTLLLAIVPIIAIAGVYEMKLSGQALKDKKEL 892
QY 893 EGAGKIATEAIENPTVVSITROKPEYMYAQSLQVYPYRNSLRKAHIFGVSFSEITQAMMY 952
Db 893 EGAGKIATEAIENPTVVSITROKPEYMYAQSLQVYPYRNSLRKAHIFGVSFSEITQAMMY 952
QY 953 FSYAGCFRFGAYLVANEFMFQDVLVFSALVFGAMAVGVSVFADYAKAKYSAAHVTM 1012
Db 953 FSYAGCFRFGAYLVANEFMFQDVLVFSALVFGAMAVGVSVFADYAKAKYSAAHVTM 1012
QY 1013 IIEKSPIDSYSPHGLKPNLTLEGNTVNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVG 1072
Db 1013 IIEKSPIDSYSPHGLKPNLTLEGNTVNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVG 1072
QY 1073 SSGCGKTSVOLLERFVDPLAGSVLIDGKEIKHLNVQWLRAHGLGIYSQEPILFDCSAEN 1132
Db 1073 SSGCGKTSVOLLERFVDPLAGSVLIDGKEIKHLNVQWLRAHGLGIYSQEPILFDCSAEN 1132
QY 1133 IAYGDSRVVSHBEIMOAAKEANTHPEITLPEKYNTRVGDGKTQLSGGQKORIAIARAL 1192
Db 1133 IAYGDSRVVSHBEIMOAAKEANTHPEITLPEKYNTRVGDGKTQLSGGQKORIAIARAL 1192
QY 1193 VRQPHILLDEATSDLTSEKVVQEAALDKAREGRTCTIVIAHRLSTIQNADLIWVFQNGK 1252
Db 1193 VRQPHILLDEATSDLTSEKVVQEAALDKAREGRTCTIVIAHRLSTIQNADLIWVFQNGK 1252
QY 1253 VKEHGTQQLLAOKGIYFSMVSVOQAGAK 1280
Db 1253 VKEHGTQQLLAOKGIYFSMVSVOQAGAK 1280

Db 1250 VKEHGTQQLLAOKGIYFSMVSVOQAGTQ 1277
RESULT 8
DVMS2
N:multidrug resistance protein 2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: A30409; S70711
R:Gros, P.; Raymond, M.; Bell, J.; Housman, D.
Mol. Cell. Biol. 8, 2770-2778, 1988
A:Title: Cloning and characterization of a second member of the mouse mdr gene family
A:Reference number: A30409; MUID:88302195
A:Accession: A30409
A:Molecule type: mRNA
A:Residues: 1-1276 <HSU>
A:Cross-references: GB:J03398; NID:g199109; PIDN:AAA39516.1; PID:g387428
R:Kirschner, L.S.
Nucleic Acids Res. 24, 2829-2834, 1996
A:Title: De novo generation of simple sequence during gene amplification.
A:Reference number: S70711; MUID:96313253
A:Accession: S70711
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 43-92 <KIR>
A:Cross-references: EMBL:U46839; NID:g1228142; PIDN:AAC52722.1; PID:g1228143
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1996
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c
tructurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: mdr2
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane
F:1-637, 653-1276/Region: duplication
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:550-554/Region: nucleotide-binding motif B
F:1049-1245/Domain: ATP-binding cassette homology <ABC2>
F:1067-1074/Region: nucleotide-binding motif A (P-loop)
F:1192-1196/Region: nucleotide-binding motif B
F:88, 94/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:432/Binding site: ATP (Lys) #status predicted
F:1072/Binding site: ATP (Lys) #status predicted
Query Match 75.1%; Score 4863; DB 1; Length 1276;
Best Local Similarity 74.8%; Pred. No. 4.3e-274;
Matches 960; Conservative 140; Mismatches 170; Indels 14; Gaps 8;
QY 1 MDPEGGRKGSARK--NFWKMGKKKK--KEKKEKKPTVSTFAMFRYSNWLRLYLGVGMA 57
Db 1 MDLEARNGTARRLDGDFELGSIHQREKKKKVNLGILLTLFRYSMDQDKLFLGLTLM 60
QY 58 ATIHGAALPLMLVFCNNTDSPA-NAGTSRNKTFPVIINESITNTQHFINHLEEMTY 116
Db 58 ATAHSGGLPLMMIVFGEMTKEDVDTG--NFSLPVNFSLSLNLPGR----ILLEEMTRY 113
QY 117 AYYYSIGAGVLVAAYIQVSFWCLAAAGROILKIRKQFFHAIMRQIGWFDVHVGELNTR 176
Db 117 AYYYSIGAGVLVAAYIQVSFWCLAAAGROILKIRKQFFHAIMRQIGWFDVHVGELNTR 176
QY 177 LDDYDSKINEGIDKIGMFFQSIATFFGIVGFTGRGKLTILVILAISPVGLSAAIWA 236
Db 177 LDDYDSKINEGIDKIGMFFQSIATFFGIVGFTGRGKLTILVILAISPVGLSAAIWA 236
QY 237 ILSSTFDKELLAYAKAGAAVEEVLAAIRTVIAFGGQKKELERYKNLEPAKIGIKKAIT 296
Db 237 ILSSTFDKELLAYAKAGAAVEEVLAAIRTVIAFGGQKKELERYKNLEPAKIGIKKAIT 296
QY 297 ANISGAFAFLLIYASALAFWGTSLVLSSEVSIGQVLVFFSVLIGAFSIGQASPSIEA 356
Db 297 ANISGAFAFLLIYASALAFWGTSLVLSSEVSIGQVLVFFSVLIGAFSIGQASPSIEA 356
QY 356 ANISGAFAFLLIYASALAFWGTSLVLSSEVSIGQVLVFFSVLIGAFSIGQASPSIEA 356
Db 356 ANISGAFAFLLIYASALAFWGTSLVLSSEVSIGQVLVFFSVLIGAFSIGQASPSIEA 356

QY 357 FANARGAAYEIFKIIDNKPISIDSYSGHKKPDNIKGNLERNVHFSYPSRKEVKILKGLN 416
Db 354 FANARGAAYVFIIDNNPKIDSERGHKKPDNIKGNLERNVHFSYPSRANIKILKGLN 413
QY 417 LKVGSGQTVALVGNCGCKSTTVQLMQRLYDPTDGMVCIDGQDRTINVRHLREITGVVS 476
Db 414 LKVGSGQTVALVGNCGCKSTTVQLMQRLYDPTDGMVCIDGQDRTINVRHLREITGVVS 473
QY 477 QEPVLFAITAEINRYGRENVTMDIEIKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGG 536
Db 474 QEPVLFTTIAENRYGRGNVTMDIEIKAVKEANAYDFIMKLPQKFTDLVGRGAQLSGG 533
QY 537 QKQRTAARALVRNPKILLDEATSDLTSEAEVQVALDKARKGRTTIVIAHRLSVRN 596
Db 534 QKQRTAARALVRNPKILLDEATSDLTSEAEVQVALDKARKGRTTIVIAHRLSVRN 593
QY 597 ADVIAGFDGVIYBKGNDHLMKPKGYFKLVMTMOTRGNEIELENATGESKESDALEMS 656
Db 594 ADVIAGFDGVIYBQGSSELKMKEGYFRLVNMOTAGSOILSEFEVEVLSDEKAAGDVA 653
QY 657 PKDSSSILIKRSTPSTHAPQODRKLCTKEDLNENVPVPSWRIILKLNSTEWYPYVG 716
Db 654 P-NGWKARIPNSTYKSLKSPH-QNRLDEETNELDANVPVPSFLKVLKLNKTEWYPYVG 711
QY 717 IFCAIINGLOPAPFSIIFSRIGITFRDEDPETKRONSMFSLVLVLGIISFTFRFLOQ 776
Db 712 TVCAIANGALQAPFSIILSEMIALFPGDD- AVKQKCNMFSLVLGLVLSFTFRFLOQ 770
QY 777 FTFGKAGELTKLRYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSRLA 836
Db 771 FTFGKAGELTLRLSMAFKAMLRQDMSFDDHKNSTGALSTRLATDAAQVQATGFKLA 830
QY 837 VITONIANLGTIISLYGNOLTLILLAIVPIITAAGVEMKMLSQALKDKKELBAG 896
Db 831 LIAQNTANLGTIISLYGNOLTLILLAIVPIITAAGVEMKMLAGNAKRDKKEMEAG 890
QY 897 KIATEAIENFTVSLTRQKFEYMAQSLQVPYRNSLRKAHIFGVFSITQAMWFSYA 956
Db 891 KIATEAIENFTVSLTRQKFEYMAQSLQVPYRNSLRKAHIFGVFSITQAMWFSYA 950
QY 957 GCFRFGAYLVANFMFNQDVLVFSALVFGAMAVGVSSPAPDYAKAKYSAAHVIMIEK 1016
Db 951 GCFRFGAYLVNFMFNQDVLVFSALVFGAMAVGVSSPAPDYAKAKYSAAHVIMIEK 1010
QY 1017 SPLIDSYSGHLPNTLEGNTNEVFNYPTRPDIPVLOGLSLKYKKGOTLALVSSGC 1076
Db 1011 SPLIDSYSGHLPNTLEGNTNEVFNYPTRPDIPVLOGLSLKYKKGOTLALVSSGC 1070
QY 1077 GKSTVQLLERFYDPLAGSVLIDGKEIKHLNVQWLAHLGIVSQEPILFDCSTAEINAYG 1136
Db 1071 GKSTVQLLERFYDPLAGSVLIDGKEIKHLNVQWLAHLGIVSQEPILFDCSTAEINAYG 1130
QY 1137 DNSRVYSHEIIMAAKEANTHIEIETLPEKYNTRVGDGKQTLGGQKQRTAARALVRP 1196
Db 1131 DNSRVYDHEIIVRAAKEANTHIEIETLPEKYNTRVGDGKQTLGGQKQRTAARALVRP 1190
QY 1197 HIILLDEATSDLTSEAEVQVALDKARKGRTTIVIAHRLSVRN 1256
Db 1191 HIILLDEATSDLTSEAEVQVALDKARKGRTTIVIAHRLSVRN 1250
QY 1257 GTHOQLLAQKGIYFSMVSVQAGAK 1280
Db 1251 GTHOQLLAQKGIYFSMVSVQAGAK 1274

RESULT 9

148123

p-glycoprotein isoform III - Chinese hamster

C;Species: Cricetus griseus (Chinese hamster)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001

C;Accession: I48123

R;Endicott, J.A.; Sarangi, F.; Ling, V.

DNA Seq. 2, 89-101, 1991

A;Title: Complete cDNA sequences encoding the Chinese hamster p-glycoprotein gene fam

A;Reference number: I48121; MUID:92135896

A;Accession: I48123

A;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1281 <RES>

A:CROSS-references: GB:M60042; NID:g191168; PIDN:AAA68885.1; PID:g191169

C:Superfamily: multidrug resistance protein; ATP-binding cassette homology

C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop

F:412-606/Domain: ATP-binding cassette homology <ABC1>

F:429-436/Region: nucleotide-binding motif A (P-loop)

F:1054-1250/Domain: ATP-binding cassette homology <ABC2>

F:1071-1078/Region: nucleotide-binding motif A (P-loop)

Query Match 74.5%; Score 4822.5; DB 2; Length 1281;

Best Local Similarity 73.7%; Pred. No. 9.8e-272;

Matches 949; Conservative 148; Mismatches 174; Indels 17; Gaps 8;

QY 1 MDPEGGRKGSAAE-----NFWKMGKSKK-KEKKKKPTVSTFAMPRYNWLDRLYMLVG 54

Db 1 MDLEAARNGTARRPGTVEGDFELGSGISNQGRNKKKVNLIPLTLFRYSDWQDKLFLMLG 60

QY 55 TMAAIIHGAALPLMLVFGNMTDSFA-NAGISRNKTFPVIIINESITNNTQHFHILHEEM 113

Db 61 TMAIAHSGSLPLMIVFGEMTDKFFVNNAG---NFSLPVNFSLSMINPGR----ILEEM 113

QY 114 TTVAYYSGIGAGLVAAVQVSWFCLAAAGROILKIRKOFFHAIMRQETGWDVHDVGBEL 173

Db 114 TRYAYYSGIGGGVLVAAVQVSWFCLAAAGROILKIRKOFFHAILRQEMGWFIDKGTTEL 173

QY 174 NPLTDDVSKINEGIDKIGMEFFQSTATFTFTGIVGTFRGWKLTLVLAISPVGLSAAI 233

Db 174 NPLTDDVSKINEGIDKIGMEFFQSTATFTFTGIVGTFRGWKLTLVLAISPVGLSAAI 233

QY 234 WAKILSSFTDKELLAYAKAGAAVEELAAIRVIAFGGOKKELERNKLEAKGIGIKK 293

Db 234 WAKILSSFTDKELLAYAKAGAAVEELAAIRVIAFGGOKKELERNKLEAKGIGIKK 293

QY 294 ATANISGAPELLIYASVALAFWYGTSLVLSSEYSIGQVLFVFFSVLIGAFSIGQASPS 353

Db 294 ATANISGAPELLIYASVALAFWYGTSLVLSSEYSIGQVLFVFFSVLIGAFSIGQASPS 353

QY 354 IEAFANARAAAYEIKIDNKPISIDSYSGHKKPDNIKGNLERNVHFSYPSRKEVKILK 413

Db 354 IEAFANARAAAYEIKIDNKPISIDSYSGHKKPDNIKGNLERNVHFSYPSRKEVKILK 413

QY 414 GLNLKVSQGTVALVGNCGCKSTTVQLMQRLYDPTDGMVCIDGQDRTINVRHLREITG 473

Db 414 GLNLKVSQGTVALVGNCGCKSTTVQLMQRLYDPTDGMVCIDGQDRTINVRHLREITG 473

QY 474 VYSQEPVLFPATTIAENIRYGRGNVTMDIEIKAVKEANAYDFIMKLPNKFDTLVGERGAQL 533

Db 474 VYSQEPVLFPATTIAENIRYGRGNVTMDIEIKAVKEANAYDFIMKLPNKFDTLVGERGAQL 533

QY 534 SGGQKQRTAARALVRNPKILLDEATSDLTSEAEVQVALDKARKGRTTIVIAHRLST 593

Db 534 SGGQKQRTAARALVRNPKILLDEATSDLTSEAEVQVALDKARKGRTTIVIAHRLST 593

QY 594 VRNADVIAGFDGVIYBKGNDHLMKPKGYFKLVMTMOTRGNEIELENATGESKESDAL 653

Db 594 VRNADVIAGFDGVIYBKGNDHLMKPKGYFKLVMTMOTRGNEIELENATGESKESDAL 653

QY 654 EMSPDKSSSILIKRSTPSTHAPQODRKLCTKEDLNENVPVPSWRIILKLNSTEWYPY 712

Db 654 EMSPDKSSSILIKRSTPSTHAPQODRKLCTKEDLNENVPVPSWRIILKLNSTEWYPY 712

QY 713 FVVGIFCAIINGLOPAPFSIIFSRIGITFRDEDPETKRONSMFSLVLVLGIISFTFR 772

Db 713 FVVGIFCAIINGLOPAPFSIIFSRIGITFRDEDPETKRONSMFSLVLVLGIISFTFR 772

QY 773 FLQGTFTFGKAGELTKLRYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIG 832

Db 773 FLQGTFTFGKAGELTKLRYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIG 832

Db 772 FLQFTGKAGAILTRLSMAFKAMLRQDMSEWFDYKNTSGALSTRLATDRAQVQOGATG 831
 QY 833 SRLAVITONIANLGTGIIISLYGWLTLALLAIPTIATAIGVEMKMLSGALKDKKEL 892
 Db 832 TRALIAQANTANLGTGIIISLYGWLTLALLAIPTIATAIGVEMKMLAGNAKDKKAL 891
 QY 893 EGAGKATEAIENFRVVSULTREQFEYVMAQSLQVPPYRNSLRKAHIFGVFSFISQAFMY 952
 Db 892 EAAGKATEAIENFRVVSULTREQFEYVMAQSLQVPPYRNSLRKAHIFGVFSFISQAFMY 951
 QY 953 FSYAGCFREGAYLVANEFNFOVLLVFSATVFGAMAVQVSSFPADYAKAKVSAAHVIM 1012
 Db 952 FSYAGCFREGAYLVANEFNFOVLLVFSATVFGAMAVQVSSFPADYAKAKVSAAHVIM 1011
 QY 1013 IIEKSLIDSYSPHGLKPNLTGENTFNEVFNYPTRPDIPVQLGLSLEVKKGQTLALVG 1072
 Db 1012 LFERQPLIDSYSGEGLWPKFEGSVTFNEVFNYPTRANPNVQLGLSLEVKKGQTLALVG 1071
 QY 1073 SSCGKSTVVQLLERYFDPLAGSLVDGKEIKHLNQLVLAHLGIVSQBPILFDCSIAEN 1132
 Db 1072 SSCGKSTVVQLLERYFDPMAGTVLLDQGEAKKLNIQLVLAHLGIVSQBPILFDCSIAEN 1131
 QY 1133 IAYGDNRSVSHHEIMQAANEIHIETLPEKYNTRYVDKGTQLSGGQKORIAIARAL 1192
 Db 1132 IAYGDNRSVSDIEIVRAAKAANIHIETLPQKYKTRYVDKGTQLSGGQKORIAIARAL 1191
 QY 1193 VROPHILLDEATSDLTSEKVVQALDKAREGRTCVIAHRLSTIQNALDIVVFQNGK 1252
 Db 1192 IROPRVLLDEATSDLTSEKVVQALDKAREGRTCVIAHRLSTIQNALDIVVFQNGK 1251
 QY 1253 VKEHGTHQQLLAQGIYFMSVSOAGAK 1280
 Db 1252 VKEHGTHQQLLAQGIYFMSVNIQAGAQ 1279

RESULT 10
 S41646
 p-glycoprotein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 02-Feb-2001
 C:Accession: S41646; S22354
 R:Brown, P.C.; Thorgeirsson, S.S.; Silverman, J.A.
 Nucleic Acids Res. 21, 3885-3891, 1993
 A:Title: Cloning and regulation of the rat mdr2 gene.
 A:Reference number: S41646; MUID:93376516
 A:Accession: S41646
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1278 <BRO>
 A:Cross-references: EMBL:L15079; NID:g310192; PIDN:AAA02937.1; PID:g310193
 R:Deuchars, K.L.; Duthie, M.; Ling, V.
 Biochim. Biophys. Acta 1130, 157-165, 1992
 A:Title: Identification of distinct p-glycoprotein gene sequences in rat.
 A:Reference number: S22351; MUID:92223089
 A:Accession: S22354
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1211-1278 <DEU>
 A:Cross-references: EMBL:X61105; NID:g56892; PIDN:CAA43417.1; PID:g1334220
 C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
 C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop
 F:409-603/Domain: ATP-binding cassette homology <ABC1>
 F:426-433/Region: nucleotide-binding motif A (P-loop)
 F:1051-1247/Domain: ATP-binding cassette homology <ABC2>
 F:1068-1075/Region: nucleotide-binding motif A (P-loop)

Query Match 74.3%; Score 4809; DB 2; Length 1278;
 Best Local Similarity 73.8%; Pred. No. 5.9e-271;
 Matches 948; Conservative 147; Mismatches 176; Indels 14; Gaps 8;

QY 1 MDPPGGKRGSAEK--NFWKMK-KSKKKKKKKKPTVSTFAMFRYSNWLDRLYMLYGTMA 57
 Db 1 MDLEARNGTARRLDGDFELGSLISNQSKKKKKNLGLPLTFRYSQDWQDKLFMLLTGAM 60

QY 58 AIHGAALPLMLLVFGNMTSFA-NAGISRNKTPFPVVIINESITNTNQHFHINHEEMTTY 116
 Db 61 AIHSGSLPLMIVFGEMTDKFDVNDAG---NFSLPVNFSLMLNPGR---ILEEMTRY 113
 QY 117 AYYSGTGAGVLVAAYTOVSFWCLAAAGROILIRKQKFFHAIHQEIGWDFVDHVGELNTR 176
 Db 114 AYYSGTGLGGVLLAAYTOVSFWTLAAGROIRKIQKFFHAILROEMGWDFDKGTTELNTR 173
 QY 177 LTDDSVKINEGIDKIGMFFOSIATFTFTGIVGTRGKLTVLILASPVLGLSAAIWA 236
 Db 174 LTDDISKISEGIDKIGMFFQAIATFFAGFIVGTRGKLTVLILASPVLGLSAAIWA 233
 QY 237 ILSFTKELLAYAKAGABEVLAAITRTVAFGGQKKELEKYNKLNLEAKGIGIKAIT 296
 Db 234 ILSFTSDKELAAAYAKAGABEALGAITRTVAFGGQKKELEKYNKLNLEAKGIGIKAIS 293
 QY 297 ANISGAFLIYASYALAFWYGTSLVLSSEYSTIGQVLTFFSVLIGAFSIGOASPIEA 356
 Db 294 ANISMGTAFLIYASYALAFWYGTSLVLSSEYSTIGQVLTFFSVLIGAFSIGOASPIEA 353
 QY 357 FANARGAAYEIFKIIDNKPISIDSYSKSGHKPDNIKGNLKNVHFSYPSRKEVKILKGLN 416
 Db 354 FPNARGAAYVIFDIIDNPKIDSFSEHGKPDNIKGNLKNVHFSYPSRANIKILKGLN 413
 QY 417 LKVSQGTVALVGNSSGCKSTTVQLMORLDYDPTDGMVCIIDQDITINVRHLREITGVVS 476
 Db 414 LKVSQGTVALVGNSSGCKSTTVQLLQRLDYDPTDGMVCIIDQDITINVRHLREITGVVS 473
 QY 477 QEPVLFAITAEINIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERAQLSGG 536
 Db 474 QEPVLFAITAEINIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERAQLSGG 533
 QY 537 QKQRIARALVRNPKILLDEATSDLTSEAVVQALDKARKGRTTIVIAHRLSTVRN 596
 Db 534 QKQRIARALVRNPKILLDEATSDLTSEAVVQALDKARKGRTTIVIAHRLSTVRN 593
 QY 597 ADVTAGFDGIVVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMS 656
 Db 594 ADVTAGFDGIVVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMS 653
 QY 657 PKDSGSLIKRRSTRSHAPQGDQDKLGTG-EDINENVPVPSFWRIILKLNSTEMPVFW 715
 Db 654 P-NKGKARIFRNSYKSKSLKSSRAHQNLVDVETNELDANVPVPSFLKVLNKLNTKEMPVFW 712
 QY 716 GIFCAIINGLOPAPSIIFSIIGTIFRDEDEPDKYQNSNMFSLVFLVIGIISITFELQ 775
 Db 713 GTLCAIANGLOPAPSIIFSIIGTIFRDEDEPDKYQNSNMFSLVFLVIGIISITFELQ 771
 QY 776 GFTFGKAGEILLTKRLRYMVERSMRLQDYSWFDKNTTGTALTTRLANDAAQVKAIGSRL 835
 Db 772 GFTFGKAGEILLTKRLRYMVERSMRLQDYSWFDKNTTGTALTTRLANDAAQVKAIGSRL 831
 QY 836 AVITQNIANGLTGIIISLYGWLTLALLAIPTIATAIGVEMKMLSGALKDKKELEGA 895
 Db 832 ALIAQNTANLGTGIIISLYGWLTLALLAIPTIATAIGVEMKMLAGNAKDKKEMEA 891
 QY 896 GKIAEAIENFRVVSULTREQFEYVMAQSLQVPPYRNSLRKAHIFGVFSFISQAFMY 955
 Db 892 GKIAEAIENFRVVSULTREQFEYVMAQSLQVPPYRNSLRKAHIFGVFSFISQAFMY 951
 QY 956 ACGRFEGAYLVANEFNFOVLLVFSATVFGAMAVQVSSFPADYAKAKVSAAHVIMIE 1015
 Db 952 ACGRFEGAYLVANEFNFOVLLVFSATVFGAMAVQVSSFPADYAKAKVSAAHVIMIE 1011
 QY 1016 KSLPIDSYPHGLKPNLTGENTFNEVFNYPTRPDIPVQLGLSLEVKKGQTLALVSSG 1075
 Db 1012 KSLPIDSYPHGLKPNLTGENTFNEVFNYPTRPDIPVQLGLSLEVKKGQTLALVSSG 1071
 QY 1076 CGKSTVQLLERYFDPLAGSLVDGKEIKHLNQLVLAHLGIVSQBPILFDCSIAENIAY 1135
 Db 1072 CGKSTVQLLERYFDPLAGSLVDGKEIKHLNQLVLAHLGIVSQBPILFDCSIAENIAY 1131

QY 1136 GDSRVVSHHEIMOAAKEANIHFTETLPKYNTRVGDGKQTLSSGQKORAIARALVRO 1195
Db 1132 GDSRVVSDDELVRRAKKEANIHFTETLPQKYNTRVGDGKQTLSSGQKORAIARALVRO 1191
QY 1196 PHILLDEATSDALDTESEKVVQVQALDKAREGRTCTIVIAHRLSTIONADLIIVVFGKVKKE 1255
Db 1192 PRVLLDEATSDALDTESEKVVQVQALDKAREGRTCTIVIAHRLSTIONADLIIVVFGKVKKE 1251
QY 1256 HGTHOQLLAOKGIYFSMVSVQAGAK 1280
Db 1252 HGTHOQLLAOKGIYFSMVSVQAGTQ 1276

RESULT 11
S55692
multidrug resistance protein homolog (mdr) - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 02-Feb-2001
C:Accession: S55692
R:Castillo, G.; Shen, H.J.; Horwitz, S.B.
Biochim. Biophys. Acta 1262, 113-123, 1995
A:Title: A homologue of the mammalian multidrug resistance gene (mdr) is functionally ex
A:Reference number: S55692; MUID:95322451
A:Accession: S55692
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1287 <CAS>
A:Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; nucleotide binding; P-loop
F:420-614/Domain: ATP-binding cassette homology <ABCI>
F:437-444/Region: nucleotide-binding motif A (P-loop)
F:1062-1258/Domain: ATP-binding cassette homology <ABC2>
F:1079-1086/Region: nucleotide-binding motif A (P-loop)

Query Match 68.5%; Score 4432; DB 2; Length 1287;
Best Local Similarity 67.7%; Pred. No. 4.6e-249;
Matches 872; Conservative 172; Mismatches 210; Indels 34; Gaps 11;

QY 2 DPEGGRGSAEKFPWKGKSKKKEKKETVSTFAMFYSNWLRLYMLVGTMAIHH 61
Db 21 DPNSNK--EKGGFF--SKFKKKEKTEKPKPKGVFTMFYRSTSDKMLMFGTIASLAH 76
QY 62 GAALPLMLVFGNMTDSFANAG--ISRNKTFPVIINESITNNQHFINHLEEMTYAYY 119
Db 77 GAALPLMLVFGNMTDSFANAG--ISRNKTFPVIINESITNNQHFINHLEEMTYAYY 127
QY 120 YSGIGAGVLVAAYIQVSWFCLAGRQLIKTRKQFFHAIMQETGWFVDVHDLNRLTLD 179
Db 128 YSGIGAGVLCAYIQISFWLSAGRQIKIRSNFFHVLRAQETGWFVDVHDLNRLTLD 187
QY 180 DVSKINGIGDKIMGFQSTATFTGTFGRGKLTILVILAIQVGLSAIAWAKILS 239
Db 188 DVSKINGIGDKIAMLQSTLTIVTGIIGFKGKLTIVVWGAISPIGMSAIAWAKILS 247
QY 240 SFTDKELLAYAKAGAAVEEVLAIARTVIAFGQKKELEKRYNKNLEAKGIGIKKAITANI 299
Db 248 AFTNKLKAYAKAGAAVEEVLISRTVFAFGQKKELEKRYNKNLEAKGIGIKKAITANV 307
QY 300 SIGAFLIIYASALAFWYQTSVLSEYSIGVQLVFFSVLIGASIGQASPSIBAFAN 359
Db 308 SIGFAFLMIYAAVSLAFWYQTSVLSEYSIGVQLVFFSVLIGASIGQASPSIBAFAN 367
QY 360 ARGAAVEIFKIIDNKPSIDSYSGKHGPNKGNLEKFNHESYPSKKEVILKGLNKKV 419
Db 368 ARGAAVTFIINDQPKIDSREGKLPKIKGIDIEFKNVIFTYPSRKDIQVLKGLNKKV 427
QY 420 QSGQTVALVGNCCGKSTTVQLMQLRDLDPDGMVCDGQDIRTINVRHLREITGVVSOEP 479
Db 428 PSKGTVALVGNCCGKSTTVQLMQLRDLDPDGMVCDGQDIRTINVRHLREITGVVSOEP 487
QY 480 VLFATTAENIRYGRNVMTDETEKAVKEANAYDFIMKLPNKFDTLVGERGQALSGGQKQ 539
Db 480 VLFATTAENIRYGRNVMTDETEKAVKEANAYDFIMKLPNKFDTLVGERGQALSGGQKQ 539

Db 488 ILFDTTIADNIRYGRNVMTDETEKAVKEANAYDFIMKLPNKFDTLVGERGQALSGGQKQ 547
QY 540 RIAIARALVRNPKILLDEATSEALDTESEAVQVQALDKAREGRTCTIVIAHRLSTIONADLI 599
Db 548 RIAIARALVRNPKILLDEATSEALDTESEAVQVQALDKAREGRTCTIVIAHRLSTIONADLI 607
QY 600 IAGFDGVIIVEKGNHDELMKEGIYFKLYMTQI---RGNIELEENATGESSESDALEM 655
Db 608 IAGFDGVIIVEQSGHKLMEGGRGVYFNLTQTVETSKDEEDLETHIYEKK-----I 660
QY 656 SPRDSSSLIKRRSTRSIIHA--POGDRKLTKFKE--DLNENVPVPSFWFIRILKLNSTWEPY 712
Db 661 PVTHSHNLVRRKSSRNTIKSKVPETEDKEVDEEEKKKEGPPVPSFFKVMKLNKPEWY 720
QY 713 FVVGIFCAIINGLQPAFSIIFRIIGIFTRDEDPETK-RQNSNMFSLVLVLVGLIISFIT 771
Db 721 FVVGIVCMIINGATQPAFAIIFRIIGVFA--GPVSQMRSSSMYSLFLFALGGVFSFIT 777
QY 772 FFLQGTFCGKAGBILTKRLRYMVFRSMLRQDYSWEDDPKNTTGALTTRLANDAAQVKGAI 831
Db 778 FFLQGTFCGKAGBILTKRLRGLSFKSMLRQDYSWEDDPKNTTGALTTRLANDAAQVKGAI 837
QY 832 GSRNAVITQNIANLGTGIIISLIYQWQLTLLIIAIVPIIAIAGVVMKMLSGOALKDKKE 891
Db 838 GTRALLAQNANLGTGIIISLIYQWQLTLLIIAIVPIIAAAGLVEMKMFAGHAKDKKE 897
QY 892 LEGAGKIATEAINFTVTVSLTREQFEYWAQSLQVYPYRNLRAHIFGVFSFISITQAMM 951
Db 898 LEKAGKISTDAVLNIRTVSVSLTRERKFEAMYEKSLGEPYRNSIKKAHLHGLTYGLSQAHH 957
QY 952 YFSYAGCFR-FGAYLVANEFMFQDVLVFSALVFCAMAVQVSSFPADPYAKAKVSAAHV 1010
Db 958 VLCLWVFSVLGALVVEGLMKLDEFLVSSAIVLGAMALGQTSFAPDITTKAMISAAHI 1017
QY 1011 IMIEKSLPIDSYPHGLKPNTELEGNTFNEVFNYPTRPDIPVLQGLSLEYKKGQTLAL 1070
Db 1018 FSLLRVPQIDSVSDQGEKPKKCSGNVVPKGVNFPYTRPDITVLQGLDISVKQGETLAL 1077
QY 1071 VSSSGCGKSTTVQVLLERFDDPLAGSVLDGKEIKHLNVOWLRAHGLGVSOEPIILFDCSIA 1130
Db 1078 VSSSGCGKSTTVQVLLERFDDPLAGSVLDGKEIKHLNVOWLRAHGLGVSOEPIILFDCSIA 1137
QY 1131 ENIAYGDNKRVSHEEIMOAAKEANIHFTETLPKYNTRVGDGKQTLSSGQKORAIAR 1190
Db 1138 DNIAYGDNKRVSHEEIMOAAKEANIHFTETLPKYNTRVGDGKQTLSSGQKORAIAR 1197
QY 1191 ALVROPHILLDEATSEKVVQVQALDKAREGRTCTIVIAHRLSTIONADLIIVVFPQN 1250
Db 1198 ALIRPKILLDEATSEKVVQVQALDKAREGRTCTIVIAHRLSTIONADLIIVVFPQN 1257
QY 1251 GKVKHGHGTHOQLLAOKGIYFSMVSVQAG 1278
Db 1258 GKVKHGHGTHOQLLAOKGIYFSMVSVQAG 1285

RESULT 12

T42228

P-glycoprotein sister - rat

N:Alternate names: multidrug resistance protein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Feb-2001

C:Accession: T42228

R:Childs, S.J.; Yeh, R.L.; Hui, D.; Ling, V.

submitted to the EMBL Data Library, June 1997

A:Description: Taxol resistance mediated by the liver-specific sister gene of P-glyco

A:Reference number: 422088

A:Accession: T42228

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1321 <CHI>

A:Cross-references: EMBL:AF010597; NID:g3273483; PID:g3273484; PIDN:AAC24753.1

C:Genetics:

A:Gene: spgq

QY 1197 HILLDEATSALDTESEKVVQEALEDKAREGRTCIIVIAHRLSTIQNADLIWVFQNGKVKHEH 1256
Db 1190 PILLDEATSALDTESEKIVQDALEVARQGRTECLVIAHRLSTIQDSDVIVMIQEGKATDR 1249
QY 1257 GTHQOQLLAQKGIYFSMVSQ 1276
Db 1250 GTHEHLLMKNDLYKRLCETQ 1269

Search completed: November 6, 2002, 18:44:57
Job time : 28.063 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:30:44 ; Search time 10.4981 Seconds
(without alignments)
4724.646 Million cell updates/sec

Title: US-09-672-725C-25

Perfect score: 6473

Sequence: 1 MDPEGGRKGSARENFWKMGK.....LLAQKGIYFSMVSQVQAKR 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	5866	90.6	1280	MDR1_HUMAN	P08193 homo sapien
2	5667.5	87.6	1276	MDR1_CRIGR	P21448 cricetus
3	5638	87.1	1276	MDR3_MOUSE	P21447 mus musculus
4	5301	81.9	1276	MDR1_MOUSE	P06795 mus musculus
5	5283	81.6	1276	MDR2_CRIGR	P21449 cricetus
6	5231.5	80.8	1277	MDR1_RAT	P43245 rattus norv
7	4915.5	75.9	1279	MDR3_HUMAN	P21439 homo sapien
8	4863	75.1	1276	MDR2_MOUSE	P21440 mus musculus
9	4822.5	74.5	1281	MDR3_CRIGR	P23174 cricetus
10	4809	74.3	1278	MDR2_RAT	P08201 rattus norv
11	3302.5	51.0	1321	AB11_HUMAN	O95342 homo sapien
12	3230.5	49.9	1321	AB11_RABIT	O90033 oryctolagus
13	3213.5	49.6	1321	AB11_MOUSE	O90730 mus musculus
14	3211.5	49.6	1321	AB11_RAT	O70127 rattus norv
15	2742	42.4	1321	MDR1_CAEEL	P34712 caenorhabdi
16	2562.5	39.6	1302	MDR4_DROME	O00449 drosophila
17	2491	38.5	1302	MDR5_DROME	O00748 drosophila
18	2341.5	36.2	1254	MDR3_CAEEL	P34713 caenorhabdi
19	2275	35.1	1362	PMO1_SCHPO	P36619 schizosacch
20	2075	32.1	1280	MDR1_LETEN	O06034 leishmania
21	1546.5	23.9	1419	MDR_PLAFF	P13568 plasmodium
22	1220.5	18.9	1290	STE6_YEAST	P12866 saccharomyc
23	1127	17.4	1336	MAM1_SCHPO	P78966 schizosacch
24	1024.5	15.8	1323	HST6_CANAL	P53706 candida alb
25	892	13.8	735	ABC8_HUMAN	O9nut2 homo sapien
26	826	12.8	582	MSBA_ECOLI	P27299 escherichia
27	817	12.6	587	MSBA_HAEIN	P44407 haemophilus
28	770	11.9	820	MDL2_YEAST	P33311 saccharomyc
29	764	11.8	598	Y288_THEMEA	O9wyc4 thermotoga
30	761	11.8	695	MDL1_YEAST	P33310 saccharomyc
31	759	11.7	1437	MRP5_HUMAN	O15440 homo sapien
32	753.5	11.6	1436	MRP5_RAT	O9qym0 rattus norv
33	743.5	11.5	1436	MRP5_MOUSE	O9rix5 mus musculus

34	743	11.5	1545	1	MRP2_HUMAN	Q92887 homo sapien
35	740.5	11.4	726	1	YFX9_SCHPO	Q9y7m7 schizosacch
36	732	11.3	1564	1	MRP2_RABIT	Q28689 oryctolagus
37	730	11.3	1325	1	MRP4_HUMAN	O15439 homo sapien
38	722	11.2	685	1	MDL1_CANAL	P37998 candida alb
39	719.5	11.1	703	1	TAP2_RAT	P36372 rattus norv
40	709	11.0	1541	1	MRP2_RAT	Q63120 rattus norv
41	708	10.9	584	1	LMRA_LACLA	O9chl8 lactococcus
42	706	10.9	575	1	YWJA_BACSU	P45861 bacillus su
43	704	10.9	604	1	YFIC_BACSU	P54719 bacillus su
44	703.5	10.9	607	1	HEPA_ANASP	P22638 anabaena sp
45	696	10.8	584	1	LMRA_LACLC	P97046 lactococcus

ALIGNMENTS

RESULT 1
MDR1_HUMAN
ID MDR1_HUMAN STANDARD; PRT; 1280 AA.
AC P08193; Q12755; Q14812;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein 1).
GN ABCB1 OR PGL1 OR MDR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8702830; PubMed=2876781;
RA Chen C.-J., Chin J.E., Ueda K., Clark D.P., Pastan I., Gottesman M.M.,
RA Roninson I.B.;
RT "Internal duplication and homology with bacterial transport proteins
RT in the mdr1 (P-glycoprotein) gene from multidrug-resistant human
RT cells.";
RL Cell 47:381-389(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90094448; PubMed=1967175;
RA Chen C.-J., Clark D.P., Ueda K., Pastan I., Gottesman M.M.,
RA Roninson I.B.;
RT "Genomic organization of the human multidrug resistance (MDR1) gene
RT and origin of P-glycoproteins.";
RL J. Biol. Chem. 265:506-514(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97190336; PubMed=9038218;
RA Chen G., Duran G.E., Steger K.A., Lacayo N.J., Jaffrezou J.P.,
RA Dumontet C., Sikic B.I.;
RT "Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein,
RT altered phenotype, and resistance to cyclosporins.";
RL J. Biol. Chem. 272:5974-5982(1997).
RN [4]
RP SEQUENCE OF 1-234 FROM N.A.
RA Smith A., Beck C., Gibson A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 178-215 AND 800-856 FROM N.A.
RX MEDLINE=90290529; PubMed=1972623;
RA Gekeler V., Weger S., Probst H.;
RT "mdr1/P-glycoprotein gene segments analyzed from various human
RT leukemic cell lines exhibiting different multidrug resistance
RT profiles.";
RL Biochem. Biophys. Res. Commun. 169:796-802(1990).
RN [6]
RP SEQUENCE OF 1-23 FROM N.A.
RA Kioka N., Tsubota J., Kakehi Y., Komano T., Gottesman M.M.,
RA Pastan I., Uedak K.;
RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED

CC DR EMBL; M14758; AAA59575.1; -;
CC DR EMBL; M29447; AAA59576.1; -;
CC DR EMBL; M29444; AAA59576.1; JOINED.
CC DR EMBL; M29424; AAA59576.1; JOINED.
CC DR EMBL; M29425; AAA59576.1; JOINED.
CC DR EMBL; M29426; AAA59576.1; JOINED.
CC DR EMBL; M29427; AAA59576.1; JOINED.
CC DR EMBL; M29428; AAA59576.1; JOINED.
CC DR EMBL; M29429; AAA59576.1; JOINED.
CC DR EMBL; M29430; AAA59576.1; JOINED.
CC DR EMBL; M29431; AAA59576.1; JOINED.
CC DR EMBL; M29432; AAA59576.1; JOINED.
CC DR EMBL; M29433; AAA59576.1; JOINED.
CC DR EMBL; M29434; AAA59576.1; JOINED.
CC DR EMBL; M29435; AAA59576.1; JOINED.
CC DR EMBL; M29436; AAA59576.1; JOINED.
CC DR EMBL; M29437; AAA59576.1; JOINED.
CC DR EMBL; M29438; AAA59576.1; JOINED.
CC DR EMBL; M29439; AAA59576.1; JOINED.
CC DR EMBL; M29440; AAA59576.1; JOINED.
CC DR EMBL; M29441; AAA59576.1; JOINED.
CC DR EMBL; M29442; AAA59576.1; JOINED.
CC DR EMBL; M29443; AAA59576.1; JOINED.
CC DR EMBL; M29444; AAA59576.1; JOINED.
CC DR EMBL; M29445; AAA59576.1; JOINED.
CC DR EMBL; M29446; AAA59576.1; JOINED.
CC DR EMBL; AF016535; AB069423.1; -;
CC DR EMBL; AC002457; AAC82531.1; -;
CC DR EMBL; M37724; AAA88047.1; -;
CC DR EMBL; M37725; AAA88048.1; -;
CC DR EMBL; X58723; CAA41558.1; -;
CC DR PIR; A25059; DVH01.
CC DR PIR; A34914; A34914.
CC DR MIM; 171050;
CC DR InterPro; IPR003593; AAA.
CC DR InterPro; IPR001140; ABC_transporter_tmem.
CC DR InterPro; IPR003439; ABC_transportr.
CC DR InterPro; IPR001687; ATP_GTP_A.
CC DR Pfam; PF00664; ABC_membrane; 2.
CC DR Pfam; PF00005; ABC_tran; 2.
CC DR SMART; SM00382; AAA; 2.
CC DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
CC KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
CC KW Multigene family.
CC FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 52 72 POTENTIAL.
CC FT TRANSMEM 120 140 POTENTIAL.
CC FT TRANSMEM 189 209 POTENTIAL.
CC FT TRANSMEM 216 236 POTENTIAL.
CC FT TRANSMEM 297 317 POTENTIAL.
CC FT TRANSMEM 326 346 POTENTIAL.
CC FT TRANSMEM 347 710 POTENTIAL.
CC FT DOMAIN 711 731 POTENTIAL.
CC FT TRANSMEM 757 777 POTENTIAL.
CC FT TRANSMEM 833 853 POTENTIAL.
CC FT TRANSMEM 854 874 POTENTIAL.
CC FT TRANSMEM 937 957 POTENTIAL.
CC FT TRANSMEM 974 994 POTENTIAL.
CC FT DOMAIN 995 1280 CYTOPLASMIC (POTENTIAL).
CC FT NP_BIND 427 434 ATP (BY SIMILARITY).

FT NP_BIND 1070 1077 ATP (BY SIMILARITY).
FT REPEAT 1 637
FT REPEAT 638 1280
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 23 23 S -> R (IN REF. 6).
FT CONFLICT 185 185 G -> V (IN REF. 1 AND 3).
FT CONFLICT 336 336 MISSING (IN REF. 3).
FT CONFLICT 412 412 G -> A (IN REF. 3).
FT CONFLICT 438 438 Q -> S (IN REF. 3).
SQ SEQUENCE 1280 AA; 141462 MW; ABC279531F43675 CRC64;

Query Match 90.6%; Score 5866; DB 1; Length 1280;
Best Local Similarity 90.8%; Pred. No. 0;
Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;

Qy 1 MDPEGRKGSA-EKNFWMGKSKKKKKKPTVSFAMFYSNWLDRMLVGTMAAI 59
Db 1 MDLEGDRNGAKKNFFKLNKS-EKDKKKKPTVSFMSFYSNWLDRMLVGTMAAI 59

Qy 60 IHGAALPLMLVFGNMTDSFANAGISRNKTPPVIIINISITNNTQHFINHLEEMTTVAY 119
Db 60 IHGAGLPLMLVFGEMTDFANAG-NLEDMSNTNRSINDTGFEN-LEEDMTRVAY 117

Qy 120 YSGIGASVLAAYIQVSFWCLAAAGQILKIRKQFFHAIMRQEIQWFDVHVGELNRLTD 179
Db 118 YSGIGAGVLAAYIQVSFWCLAAAGQILKIRKQFFHAIMRQEIQWFDVHVGELNRLTD 177

Qy 180 DYSKINEGIDKIGMFFOSIATFTFTGTVGTGRGKLTVLILAI SPVLGSLAAWAKILS 239
Db 178 DYSKINEGIDKIGMFFOSMATFTFTGTVGTGRGKLTVLILAI SPVLGSLAAWAKILS 237

Qy 240 SFTDKELLAYAKAGAAVEVLAAITVIAFGQKKELEKRYNKNLEAKGIGIKKAITANI 299
Db 238 SFTDKELLAYAKAGAAVEVLAAITVIAFGQKKELEKRYNKNLEAKGIGIKKAITANI 297

Qy 300 SIGAAFLLIYASALAFWYGTSLVLSSEYSIGQVLTVEFVSIGVLTGAFSGASPIEFAN 359
Db 298 SIGAAFLLIYASALAFWYGTTLVLSGEYSIGQVLTVEFVSIGVLTGAFSGASPIEFAN 357

Qy 360 ARGAAEYEFKIIDNKPSIDSYSKSGHKPDNLKGNLFKNVHFSYPSRKEVKILGLNLKV 419
Db 358 ARGAAEYEFKIIDNKPSIDSYSKSGHKPDNLKGNLFKNVHFSYPSRKEVKILGLNLKV 417

Qy 420 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTDMQVCIQDQDITINVRHLREITGVVSOEP 479
Db 418 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTDMQVCIQDQDITINVRHLREITGVVSOEP 477

Qy 480 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPKFDTLVGERAQLSGGQK 539
Db 478 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPKFDTLVGERAQLSGGQK 537

Qy 540 RTAIALRVNPKILLDEATSDTESAVVQVALDKARKGRTTIVIAHRLSTVRNADV 599
Db 538 RTAIALRVNPKILLDEATSDTESAVVQVALDKARKGRTTIVIAHRLSTVRNADV 597

Qy 600 TAGFDDGVIVEKGNHDELMKEKGIYFKLVMTQTRGNETELENAATGESKESDALEMSPKD 659
Db 598 TAGFDDGVIVEKGNHDELMKEKGIYFKLVMTQTRGNETELENAATGESKESDALEMSPKD 657

Qy 660 SGSSLIKRSTRSRTHAPQGDQRKLTGKTEDLNENVPVPSFWRILKLNSTFWVVGIFC 719
Db 658 SRSSLIRKSTRSRVRSQAQDKLSTKEALDESIPPVSFWRIMKLNTEWPIYFVVGIFC 717

Qy 720 AILINGLOPAFSIIFSRIGITRDEDPETKRONSMFSLVFLVLIIGIISFTFFLQGF 779
Db 718 AILINGLOPAFAIIEFKIIGVTFTRIDDPETKRONSLFLLFLALGIISFTFFLQGF 777

Qy 780 GKAGEILTRLRYMVFRLSRMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGIGSLAVIT 839
Db 778 GKAGEILTRLRYMVFRLSRMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGIGSLAVIT 837

QY 840 QNIAIAGTGIISLIYQWLTLLLAIVPIIAIAGVWEMKMLSGALKDKKELEGAGKIA 899
 Db 838 QNIAIAGTGIISLIYQWLTLLLAIVPIIAIAGVWEMKMLSGALKDKKELEGAGKIA 897
 QY 900 TEAIEFRTVSLTREQFEYMYAOSLQVPYRNSLRKAHIFGVSPISITQAMMYFSYAGCF 959
 Db 898 TEAIEFRTVSLTREQFEYMYAOSLQVPYRNSLRKAHIFGVSPISITQAMMYFSYAGCF 957
 QY 960 REGAYLVANEFNFODVLLVFAIYVFGAMVGOVSSFPADYAKAKYAAHVIMIIKESPL 1019
 Db 958 REGAYLVANEFNFODVLLVFAIYVFGAMVGOVSSFPADYAKAKYAAHVIMIIKESPL 1017
 QY 1020 IDSYSPHGLKPNLTENYTFEVEFNYPTRDPVLPVQGLSLVKKGTQALVSGSGCGKS 1079
 Db 1018 IDSYSTEGMLPNTLEGNVTFEVEFNYPTRDPVLPVQGLSLVKKGTQALVSGSGCGKS 1077
 QY 1080 TVVQLLEFYDPLAGSLVDGKEIKHLNVQWLRHAHLGTVSOPILFDCSIAENIAYGDS 1139
 Db 1078 TVVQLLEFYDPLAGSLVDGKEIKHLNVQWLRHAHLGTVSOPILFDCSIAENIAYGDS 1137
 QY 1140 RVVSHEEIMOAKEANIHHFTETLPEKYNTVRGDKGTOLSGGOKORIAIARALYRPHIL 1199
 Db 1138 RVVSHEEIMOAKEANIHHFTETLPEKYNTVRGDKGTOLSGGOKORIAIARALYRPHIL 1197
 QY 1200 LDEATSAIDTESEKVVQALDKAREGRTICIVIAHRLSTIONADLIYVFQNGKVEHGT 1259
 Db 1198 LDEATSAIDTESEKVVQALDKAREGRTICIVIAHRLSTIONADLIYVFQNGKVEHGT 1257
 QY 1260 QOLLAQKGIYFSMVSQVQAKR 1281
 Db 1258 QOLLAQKGIYFSMVSQVQAKR 1279

RESULT 2
 MDRL_CRIGR
 ID MDRL_CRIGR STANDARD; PRT; 1276 AA.
 AC P21448;
 DT 01-NOV-1991 (Rel. 18, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Multidrug resistance protein 1 (P-glycoprotein 1).
 GN ABCB1 OR PGY1 OR PGPI.
 OS Cricetulus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92135896; PubMed=1685679;
 RA Endicott J.A., Sarangi F., Ling V.;
 RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
 gene family";
 RL DNA Seq. 2:89-101(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91154265; PubMed=1671863;
 RA Devine S.E., Hussain A., Davide J.P., Melera P.W.;
 RT "Full length and alternatively spliced pgp1 transcripts in multidrug-
 resistant Chinese hamster lung cells";
 RL J. Biol. Chem. 266:4545-4555(1991).
 RN [3]
 RP SEQUENCE OF 706-1276 FROM N.A.
 RX MEDLINE=88122132; PubMed=2893255;
 RA Endicott J.A., Juranka P.F., Sarangi F., Gerlach J.H., Deuchars K.L.,
 RA Ling V.;
 RT "Simultaneous expression of two P-glycoprotein genes in
 drug-sensitive Chinese hamster ovary cells.";
 RL Mol. Cell. Biol. 7:4075-4081(1987).
 CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
 CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT

CC CAPABILITIES: PGPI AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
 CC CANNOT.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M60040; AAA68883.1; -;
 CC EMBL; M59253; AAA37004.1; -;
 CC EMBL; M17897; AAA37006.1; -;
 CC PIR; A38696; DVHYIC.
 CC InterPro; IPR003593; AAA.
 CC InterPro; IPR001140; ABC_transporter_tmemb.
 CC InterPro; IPR003439; ABC_transport.
 CC InterPro; IPR001687; ATP_GTP_A.
 CC Pfam; PF00664; ABC_membrane; 2.
 CC Pfam; PF00005; ABC_tran; 2.
 CC SMART; SM00382; AAA; 2.
 CC PROSITE; PS00211; ABC_TRANSPORTER; 2.
 CC ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
 KW Multigene family.
 KW DOMAIN 1 50 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 213 233 POTENTIAL.
 FT TRANSMEM 294 314 POTENTIAL.
 FT TRANSMEM 323 343 POTENTIAL.
 FT DOMAIN 344 707 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 708 728 POTENTIAL.
 FT TRANSMEM 754 774 POTENTIAL.
 FT TRANSMEM 830 850 POTENTIAL.
 FT TRANSMEM 851 871 POTENTIAL.
 FT TRANSMEM 934 954 POTENTIAL.
 FT TRANSMEM 971 991 POTENTIAL.
 FT DOMAIN 992 1276 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 424 431 ATP (POTENTIAL).
 FT NP_BIND 1067 1074 ATP (POTENTIAL).
 FT REPEAT 1 635
 FT REPEAT 636 1276
 FT CONFLICT 338 339 GA -> AP (IN REF. 2).
 SQ SEQUENCE 1276 AA; 140925 MW; 443F92A186B4DFF CRC64;
 Query Match 87.6%; Score 5667.5; DB 1; Length 1276;
 Best Local Similarity 87.1%; Pred. No. 3.3e-316;
 Matches 1116; Conservative 77; Mismatches 82; Indels 7; Gaps 3;
 QY 1 MDPEGRKGSAAENFWMKKKKKKKKPTVSTFAMFPRYSNWLDRMLVGLTMAAII 60
 Db 1 MEFEEDFSGRKDKNFKMGKRS-KKBBKPKPVSVFTMFYAGWLDRLVGLTAAII 59
 QY 61 HGAALPLMLVFCNMTDSPANAG-ISRNTKFPVINESITNTQHFHINHEEMTYAY 119
 Db 60 HGVALPLMLVFGMDTDSFASVGNIPNTAT-----NNATQVNASDFGKLEEMTYAY 114
 QY 120 YSGIGAGVLVAAYIQVSWFCWLAAGRQLKIRKOFFHAIMRQETGWFVDHVGELNRLD 179
 Db 115 YTGIGAGVLIVAYIQVSWFCWLAAGRQIHKIRQKFFHAIMNQEIWFVDHVGELNRLD 174
 QY 180 DVSKINEGIDGKTMGPFQSIATFTFTGIVGTRGWKLTILVLAISVPLGSLRAIWAII 239
 Db 175 DVSKINEGIDGKTMGPFQSIATFTFTGIVGTRGWKLTILVLAISVPLGSLRAIWAII 234
 QY 240 SFTDKELAYAKAGAAVEEVLAIKRTVIAFGGOKKELRYNKNLEAKGIGKAITANI 299
 Db 235 SFTDKELQAYAKAGAAVEEVLAIKRTVIAFGGOKKELRYNKNLEAKGIGKAITANI 294
 QY 300 SIGAAFLIIYASVALAFWYGTSLVLSSEYSIGQVLTFTVSVLIGAFSIGQSPSTEAFAN 359

```

Db 295 SMGAPELLIYASALAFWGTSLVTSKEYSIGQVLTVFAVLIGAFSGQASPNIEAFAN 354
QY 360 ARGAAYEIKIIDNKPSIDYSKSGHKPDNIKGNLEFKNVHFSYPSRKEVKILKGLNLKV 419
Db 355 ARGAAYEINIDNKPSIDFSKNGYKPDNIKGNLEFKNVHFSYPSRQDVQLKGLNLKV 414
QY 420 QSGQTVGVNSGCGKSTTVQMLQRLYDPTDGMVCIDGQDRTINVRHLREITGVVSQEP 479
Db 415 QSGQTVGVNSGCGKSTTVQMLQRLYDPTDGMVCIDGQDRTINVRHLREITGVVSQEP 474
QY 480 VLPATTIAENIRYGRNVTWDEIEKAVKANAYDFIMKLPNKFDTLVGRGAQLSGGQKQ 539
Db 475 VLPATTIAENIRYGRNVTWDEIEKAVKANAYDFIMKLPNKFDTLVGRGAQLSGGQKQ 534
QY 540 RIAIARALYRNPKILLIDENATSDALDESEAVQVALDKARKGRTTIVIAHRLSTVRNADV 599
Db 535 RIAIARALYRNPKILLIDENATSDALDESEAVQVALDKARKGRTTIVIAHRLSTVRNADI 594
QY 600 IAGFDGVIKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSKPD 659
Db 595 IAGFDGVIKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSKPD 654
QY 660 SSSLLIKRSTRRSIHAPQOGQKLTGKEDLNENPPVFWFRIKLKLNSTEWYFVVGIFC 719
Db 655 SSSLLIKRSTRRSIHAPQOGQKLTGKEDLNENPPVFWFRIKLKLNSTEWYFVVGIFC 714
QY 720 AINGLOPAFSTIFSRITGIFTRDEDPETKRONSMFVFLVILGLIGSIFTFELOGFTF 779
Db 715 AINGLOPAFSTIFSRITGIFTRDEDPETKRONSMFVFLVILGLIGSIFTFELOGFTF 774
QY 780 GKAGEILTRRLRYMFRSLRQDVSNFDDPKNTGALTTLRLANDAAQVKGAGISLAVIT 839
Db 775 GKAGEILTRRLRYMFRSLRQDVSNFDDPKNTGALTTLRLANDAAQVKGAGISLAVIT 834
QY 840 QNTANLGTGIIISLIYGWOLTLTLLLAIVPIIAIAGVVENKMLSGQALDKKELEGAKRTA 899
Db 835 QNTANLGTGIIISLIYGWOLTLTLLLAIVPIIAIAGVVENKMLSGQALDKKELEGAKRTA 894
QY 900 TEAIENFRTVSLTROKPEYMYAQSLQVYRNSLRKAHIFGVFSITQAMMYFSYAGCF 959
Db 895 TEAIENFRTVSLTROKPEYMYAQSLQVYRNSLRKAHIFGVFSITQAMMYFSYAGCF 954
QY 960 RFQAYLVANEFMNFQDVLVFSIAIVFGAMAGVQVSSFPADYAKAKVSAHVIMIEKSP 1019
Db 955 RFQAYLVANEFMNFQDVLVFSIAIVFGAMAGVQVSSFPADYAKAKVSAHVIMIEKSP 1014
QY 1020 IDSYSPLKPNLTGNTVFNENYVFNPTRPDIPVLOGLSLEVKKGOTLALVSSGCGKS 1079
Db 1015 IDSYSPLKPNLTGNTVFNENYVFNPTRPDIPVLOGLSLEVKKGOTLALVSSGCGKS 1074
QY 1080 TVVQLLERFYDPLAGSVLIDGKEIKHLNQLWLAHILGIVSQEPILFDCSIAENIAYGDS 1139
Db 1075 TVVQLLERFYDPLAGSVLIDGKEIKHLNQLWLAHILGIVSQEPILFDCSIAENIAYGDS 1134
QY 1140 RVYSHEIMQAQKANTHIEPLPKYNTFRVGDKTQISGGQKQRIARALVROPHIL 1199
Db 1135 RVYSQDEIERAAKEANTHIEPLPKYNTFRVGDKTQISGGQKQRIARALVROPHIL 1194
QY 1200 LLEENATSDALDESEAVQVALDKARKGRTTIVIAHRLSTONADLLIVFQNGKVEHGH 1259
Db 1195 LLEENATSDALDESEAVQVALDKARKGRTTIVIAHRLSTONADLLIVFQNGKVEHGH 1254
QY 1260 QLLAQAQGIYFVMSVQAGAKR 1281
Db 1255 QLLAQAQGIYFVMSVQAGAKR 1276

RESULT 3
MDR3_MOUSE STANDARD; PRT; 1276 AA.
AC P21447;
DT 01-MAY-1991 (Rel. 18, Created)
```

```

DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (P-glycoprotein 3) (MDR1A).
GN ABCB4 OR PGY3 OR PGY-3 OR MDR3 OR MDR1A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205845; PubMed=1969610;
RA Devault A., Gros P.;
RT "Two members of the mouse mdr gene family confer multidrug resistance
with overlapping but distinct drug specificities.";
RL Mol. Cell. Biol. 10:1652-1663(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90287150; PubMed=1972547;
RA Hsu S.I.H., Cohen D., Kirschner L.S., Lothstein L., Hartstein M.,
Horwitz S.B.;
RT "Structural analysis of the mouse mdr1a (P-glycoprotein) promoter
reveals the basis for differential transcript heterogeneity in
multidrug-resistant J774.2 cells.";
RL Mol. Cell. Biol. 10:3596-3606(1990).
RN [3]
RP SEQUENCE OF 173-1276 FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=93308614; PubMed=2473069;
RA Hsu S.I.H., Lothstein L., Horwitz S.B.;
RT "Differential overexpression of three mdr gene family members in
multidrug-resistant J774.2 mouse cells. Evidence that distinct P-
glycoprotein precursors are encoded by unique mdr genes.";
RL J. Biol. Chem. 264:12053-12062(1989).
CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
RELATED BUT DISTINCT CELLULAR GENES.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M30697; AAA39517.1; -
DR EMBL; M33581; AAA39514.1; -
DR EMBL; M33580; AAA39518.1; -
DR EMBL; M24417; AAA03243.1; -
DR PIR; A34175; DVMS1A.
DR PIR; A34786; A34786.
DR MGD; MGI-97570; Abcb4.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
Multigene family.
FT DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 48 71 POTENTIAL.
FT TRANSMEM 117 136 POTENTIAL.
FT TRANSMEM 188 205 POTENTIAL.
FT TRANSMEM 212 232 POTENTIAL.
FT TRANSMEM 298 318 POTENTIAL.
FT TRANSMEM 327 346 POTENTIAL.
FT DOMAIN 347 707 CYTOPLASMIC (POTENTIAL).
```


Mol. Cell. Biol. 10:6036-6040(1990).
 CC - FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
 CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
 CC - SUBCELLULAR LOCATION: Integral membrane protein.
 CC - PTM: SEVERAL PHOSPHORYLATED SERINE RESIDUES ARE PRESENT IN THE
 CC LINKER DOMAIN.
 CC - MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
 CC RELATED BUT DISTINCT CELLULAR GENES.
 CC - SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; M14757; AAA79005.1; -;
 DR EMBL; M60348; AAA39513.1; -;
 DR PIR; A33719; DVMS1.
 DR MGD; MGI:97568; Abcb1.
 DR InterPro; IPR003593; AAA.
 DR InterPro; IPR001140; ABC_transporter_tmern.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR001687; ATP_GTP_A.
 DR Pfam; PF00664; ABC_membrane; 2.
 DR Pfam; PF00005; ABC_tran; 2.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
 KW Multigene family; Phosphorylation.
 FT DOMAIN 1 47
 FT TRANSMEM 48 71
 FT TRANSMEM 119 139
 FT TRANSMEM 199 219
 FT TRANSMEM 296 316
 FT TRANSMEM 327 347
 FT TRANSMEM 348 379
 FT DOMAIN 348 705
 FT TRANSMEM 710 730
 FT TRANSMEM 755 775
 FT TRANSMEM 831 851
 FT TRANSMEM 855 882
 FT TRANSMEM 935 955
 FT TRANSMEM 966 986
 FT DOMAIN 987 1276
 FT NP_BIND 426 433
 FT NP_BIND 1068 1075
 FT REPEAT 1 635
 FT REPEAT 636 1276
 FT CARBOHYD 73 73
 FT CARBOHYD 91 91
 FT CARBOHYD 96 96
 FT CARBOHYD 103 103
 FT CARBOHYD 1276 AA; 140993 MW; 1804D0F011B0FF4E CRC64;
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 ATP (BY SIMILARITY).
 ATP (BY SIMILARITY).

Query Match 81.9%; Score 5301; DB 1; Length 1276;
 Best Local Similarity 80.3%; Pred. No. 2.9e-295;
 Matches 1030; Conservative 130; Mismatches 114; Indels 8; Gaps 5;

QY 1 MDEGGRKSAENFWKGGKSKKKEKKKPTVTFAMFRYSNWLRLMLVGTMAII 60
 DB 1 MEFEENLKGRAADNFKSMGKKS-KKEKKKKKPAVGFMFRYADWLKLCMLTGLTAAII 59
 QY 61 HGAAFLPMLVFGNMTDSFANAGISRNKTFPVITNESITNNQHEIN-HLEEMTYAY 119
 DB 60 HGTLLPLMLVFGNMTDSFKAAS---ILPSITNQSNGSNSTLIISNSLEEMAIYAY 116
 QY 120 YSGIAGVLVAAYIQVSWCLAAAGROILKIQKOFHAIMRQIGWFDVHVGELNRLTD 179
 DB 117 YTGIGAGVLIVAYIQVSWCLAAAGROILKIQKOFHAIMRQIGWFDVHVGELNRLTD 176
 QY 180 DVSKINGEGDKIGMFFQSIATFFTGIVFGTRGKWLIVILAIAPVLGLSAAIWAII 239

177 DVSKINDIGDKIMFFQSIATFFLAGIIFGTSWGLTIVILAVSPILGLSALWAKVLT 236
 QY 240 SFTDKELLAYAKAGAAEEVLAIRVIAFGGOKKELRYNKNLEAKGIGIKKAITANI 299
 DB 237 SFTNKLQAYAKAGAAEEVLAIRVIAFGGOKKELRYNKNLEAKGIGIKKAITASI 296
 QY 300 SIGAAFLIIVASYALAFWGTSTLVLSSEYSGIQLVTFVSVLIGAFSIOGASPSIAFAN 359
 DB 297 SIGAYLVVASYALAFWGTSTLVLSSEYSGIQLVTFVSVLIGAFSIOGASPSIAFAN 356
 QY 360 ARGAAVEIKIINKPSIDSYSGSKHKNKGNLEFKNVHSPSKVEVKILGLNLKV 419
 DB 357 ARGAAVEIKIINKPSIDSYSGSKHKNKGNLEFKNVHSPSKVEVKILGLNLKV 416
 QY 420 QSGOTVALVNSCGSGSTTVQLMQLVDPDGMVCIDGQDIRTINVRHLREITGVVSQEP 479
 DB 417 KSGOTVALVNSCGSGSTTVQLMQLVDPDGMVCIDGQDIRTINVRHLREITGVVSQEP 476
 QY 480 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKEDTLVGERGAOLSGGOK 539
 DB 477 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKEDTLVGERGAOLSGGOK 536
 QY 540 RIAIARALVNPKNLLDDEATSEAVQVADLKAKGRTTIVIAHRLSTVRNADY 599
 DB 537 RIAIARALVNPKNLLDDEATSEAVQVADLKAKGRTTIVIAHRLSTVRNADY 596
 QY 600 IAGFDGIVVEKGNHDELMKEGIYFKLVMTQTRGNEIELEENATGESKESDALESPKD 659
 DB 597 IAGFDGIVVEKGNHDELMKEGIYFKLVMTQTRGNEIELEENATGESKESDALESPKD 656
 QY 660 SGLSLIKRRSTRSIHAPQGDQRKLGTEDLNENVPVSWFRLKLNSTEWPFVVGIFC 719
 DB 657 SKSPLI--RRSIYRSVHRKQDQERRLSMKEAVDEVDPLVSEFWRLNLSNWPVLLVGLVC 715
 QY 720 AILINGLOPAFSIIFSRIGIETREDPEKTRONSNMFSVLFVLVGLTIFITPEFLOGTF 779
 DB 716 AVINGCQPVFAIVFSRIVGVSRDDHDETRKONCNLFSLFVLLVGLISEVTFYFLOGTF 775
 QY 780 GKAGEILTCLRVMVFRSMLRQDVSFDPKNTGTALTTTLANDAAQVKGAIQSLRAVIT 839
 DB 776 GKAGEILTCLRVMVFRSMLRQDVSFDPKNTGTALTTTLANDAAQVKGAIQSLRAVIT 835
 QY 840 QNTANLGTGIIISLYGWOLTLALLAIVPIIATAGVEMKLSQALKDKKKELEGAKIA 899
 DB 836 QNVANLGTGIIISLYGWOLTLALLAIVPIIATAGVEMKLSQALKDKKKELEGAKIA 895
 QY 900 TEALENERTVSLTREQKFEYMYAQSLQVPRNLSRAHIFGVSFSTQAMMYFSYAGCF 959
 DB 896 TEALENERTVSLTREQKFEYMYAQSLQVPRNLSRAHIFGVSFSTQAMMYFSYAGCF 955
 QY 960 RFGAYLVANEFMNFQDVLVFSALVFGAMVQVSSFPADYAKAKVSAHVIMIEKSP 1019
 DB 956 RFGAYLVANEFMNFQDVLVFSALVFGAMVQVSSFPADYAKAKVSAHVIMIEKSP 1015
 QY 1020 IDSYSPIGLKPTNLEGNVTFNEVFNPTRPDIPVLQGLSLEVKKGOTLALVSSGCGKS 1079
 DB 1016 IDSYSPIGLKPTNLEGNVTFNEVFNPTRPDIPVLQGLSLEVKKGOTLALVSSGCGKS 1075
 QY 1080 TVQOLLERFDPDLAGSVLIDGKEIKHLNVQWLAHLRAHIGIVSQEPILFDCSTAEINAYG 1139
 DB 1076 TVQOLLERFDPDLAGSVLIDGKEIKHLNVQWLAHLRAHIGIVSQEPILFDCSTAEINAYG 1135
 QY 1140 RVVSHETIMOAKDANIHHTIETLPEKYNTRVGDGKQTLGGGQKQRIATARALVRPHIL 1199
 DB 1136 RAVSHEETVRAAKKANIHQFIDSLPDKNYTRVGDGKQTLGGGQKQRIATARALVRPHIL 1195
 QY 1200 LLDDEATSEKVVQVEALDKAREGTCIVIAHRLSTIONADLIVVFONGVKVKEGHT 1259
 DB 1196 LLDDEATSEKVVQVEALDKAREGTCIVIAHRLSTIONADLIVVFONGVKVKEGHT 1255
 QY 1260 QOLLAQRGIFYSMVSVQAGAKR 1281
 DB 1260 QOLLAQRGIFYSMVSVQAGAKR 1281

901 EAIENFRTVSLTREQKFEYMAQSLQVYPYRNSLRKAHIFGVFSFISQAMMYFSYAGCFR 960
 Db 897 EAIENFRTVSLTREQKFNMYAQSILQIPYRNALKAHVFGITFSFTQAMMYFSYAACFR 956
 QY 961 FGAYLVANFEMFQDVLVFSIAIVGAMAVGOVSFADPYAKAKYSAAHVIMIEKSPLI 1020
 Db 957 FGAYLVAHQIMTFENVMVFSIAVFGAIAAGNASSFADPYAKAKYSAASHIRIMEKIPSI 1016
 QY 1021 DSVSPHGLKPNLTLEGNTVENVVFNYPTRPDIPVQLQGLSLEVKKGOTLALVSSGCGKST 1080
 Db 1017 DSVSTGLAPNLEGNVKNENVVFNYPTRPDIPVQLQGLSLEVKKGOTLALVSSGCGKST 1076
 QY 1081 VVOLLERFYDPLAGSVLIDGKEIKHLNVOMLRAHLGIYSQEPILFDCSIAENIAYGDNRS 1140
 Db 1077 VVOLLERFYDPMAGTVFLDGKEIKQLNVOMLRAHLGIYSQEPILFDCSIAENIAYGDNRS 1136
 QY 1141 VYSHEIMOAAKEANTHHFIEPLPEKYNTRVGDGKTQLSGGOKQRIATARALVROPHILL 1200
 Db 1137 VVSQDIERAKEANTHQPESLPDKYNTRVGDGKTQLSGGOKQRIATARALVROPHILL 1196
 QY 1201 LDEATSAIDTESEKVVQVEALDKAREGTCIVIAHRLSTIONADLIWVFQNGVKKEGTHQ 1260
 Db 1197 LDEATSAIDTESEKVVQVEALDKAREGTCIVIAHRLSTIONADLIWVFQNGVKKEGTHQ 1256
 QY 1261 QLLAQKGIYFSMVQAGAKR 1281
 Db 1257 QLLAQKGIYFSM--VQAGAKR 1275

RESULT 6
 MDRL_RAT ID MDRL_RAT STANDARD; PRT: 1277 AA.
 AC P43245;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Multidrug resistance protein 1 (P-glycoprotein 1).
 GN ABCB1 OR PGY1 OR MDR1 OR MDR1B.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92039081; PubMed=1682220;
 RA Silverman J.A., Raunio H., Gant T.W., Thorgeirsson S.S.;
 RT "Cloning and characterization of a member of the rat multidrug
 resistance (mdr) gene family.";
 RL Gene 106:229-236(1991).
 CC -I- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
 CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -I- PTM: SEVERAL PHOSPHORYLATED SERINE RESIDUES ARE PRESENT IN THE
 CC LINKER DOMAIN.
 CC -I- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
 CC RELATED BUT DISTINCT CELLULAR GENES.
 CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M01855; ; NOT_ANNOTATED_CDS.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001140; ABC_transporter_tmern.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00664; ABC_membrane; 2.
 DR Pfam: PF00005; ABC_tran; 2.

SMART: SM00382; AAA; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
 KW Multigene family.
 FT DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 48 71 POTENTIAL.
 FT TRANSMEM 119 139 POTENTIAL.
 FT TRANSMEM 199 219 POTENTIAL.
 FT TRANSMEM 296 316 POTENTIAL.
 FT TRANSMEM 327 347 POTENTIAL.
 FT DOMAIN 348 709 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 710 730 POTENTIAL.
 FT TRANSMEM 755 775 POTENTIAL.
 FT TRANSMEM 832 852 POTENTIAL.
 FT TRANSMEM 854 874 POTENTIAL.
 FT TRANSMEM 937 957 POTENTIAL.
 FT TRANSMEM 968 988 POTENTIAL.
 FT DOMAIN 989 1277 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 426 433 ATP (BY SIMILARITY).
 FT NP_BIND 1070 1077 ATP (POTENTIAL).
 FT REPEAT 1 635
 FT REPEAT 636 1277
 FT CARBOHYD 91 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 1277 AA; 141386 MW; 8AFDD619D2934C1 CRC64;
 Query Match 80.8%; Score 5231.5; DB 1; Length 1277;
 Best Local Similarity 80.0%; Pred. No. 2.7e-291;
 Matches 1029; Conservative 131; Mismatches 111; Indels 15; Gaps 9;

QY 1 MDPGGRKSGSAEKNFWMKGGKSKKKEKPTVSTFAMPFYSNWLDRMLVMTMAAII 60
 Db 1 MEFEGLNGRADKNFSKMGKSKK -EKEKPAVGIIFGMFRIADWLKLCMALGTAAII 58
 QY 61 HGAALPLMLLVFGNMTDSFANAGISRN-KTFPPVINESITNTQHFHN--HLEEMTYA 117
 Db 59 HGTLLPLMLLVFGYMTDSFTP--SRDPHSDRATNQSEINST-HTVSDTSLEEDMAYA 114
 QY 118 YYYSGIGAGVLVAAYIQVSWFCLAAAGQILKIRQFPHAIMRQFIGHDFVDVDELNRL 177
 Db 115 YYYTGIGAGVLIVAYIQVSLWCLAAAGRIKIRQFPHAIMNQFIGNDFVNDAGELNRL 174
 QY 178 TDDVSKINEGIDKTMGFQSIATFTFGFTVGRGKLTVLVLAISPVLGSLAAIWAII 237
 Db 175 TDDVSKINDGIDKLMGFFQSIITTSAGFIIGFISGKMLVLVLAVSPLGLSSAMWAKV 234
 QY 238 LSSFTDKELLAYAKAGAAVEEVLAARITVIAFGGQKKELERYNKNLEBEAKGIGIKKAITA 297
 Db 235 LTSFTNKLQAYAKAGAAVEEVLAARITVIAFGGQKKELERYNKNLEBEAKRVGKIKAITA 294
 QY 298 NISGAAPELLIYASALAFWYGTSLVLSSEYSIGQVLTVPFVSLIGAFSGQASPSIEAF 357
 Db 295 NISIGIAYLLVYASALAFWYGTSLVLSSEYSIGQVLTVPFVSLIGAFSGHAPNTEAF 354
 QY 358 ANARCAAYEIKIIDNKPSIDSYSGKHGPDNIKGNLEFKNVHFSEYSPRKEVKILKGLNL 417
 Db 355 ANARCAAYEIKIIDNEPSIDSFSTKGHKPDSIMGNLEFKNVFNYSRSEVKILKGLNL 414
 QY 418 KVQSQTVVALVGNSSCGKSTTVQLMQRLYDPTDGMVCIDGQDIRTINVRLREITGVVSQ 477
 Db 415 KVKSQTVVALVGNSSCGKSTTVQLLRLYDPIEGEVSIDGQDIRTINVRLREITGVVSQ 474
 QY 478 EPLVFTATTIANIRYGRNVMTDEIEKAVKANAYDFIMKLPNKFDTLVGRGQALSGGQ 537
 Db 475 EPLVFTATTIANIRYGRNVMTDEIEKAVKANAYDFIMKLPNKFDTLVGRGQALSGGQ 534
 QY 538 KORTAIALRVNPKLILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNA 597
 Db 535 KORTAIALRVNPKLILLDEATSDALDTESEAVVQVALDKAREGRTTIVIAHRLSTVRNA 594
 QY 598 DVIAGFDGDIVEGKGNHDELMKEKGIYFKLVMTQTRGNEIELEENATGESKESDALEMS 657
 Db 595 DVIAGFDGDIVEGQNHHELMKEKGIYFKLVMTQTRGNEIEPPGNNAVESQSDTGASELTS 654

QY	658	KDGSLLIKRRSTRBSIHAPQODRKLGTKEDLNENVPVSWFRILKLNSTWPFVVG	717
Db	655	EESKSPGL-RRSIRSRHRRQDERRLSKEDVDPMVSWFRILKLNISEWPLYVVG	713
QY	718	FCALINGGLOPAPFSIFRIGIFTRDEDPETKRONSMFSLFLVLGLGIISFITFLQGF	777
Db	714	LCAVINGCIOPFAIVEFSKIVGVFSRDDDDHETKORCNLFSLFLVMGMSFVTFQGF	773
QY	778	TFGRAGELTKRLRYWFRSMRLQDVSWDFDKNTGTALTRLANDAAQVKAIGSRUAV	837
Db	774	TFGRAGELTKRLRYWFRSMRLQDVSWDFDKNTGTALTRLANDAAQVKAIGSRUAV	833
QY	838	ITQNIANLGTGIIIS--LIYWGOLFLLALLAIVPIIAIAGVEMKMLSGQALKDKKELEGA	895
Db	834	VTQNVANLGTGIIISLVLYVGOLFLLVVIPLVGLGIIEMKLLSGQALKDKKELEIS	893
QY	896	GKIATEAIENPRTVVSLSLREQKFETMYAQSLOQVPYRNSLRKAHIFGVFSFITQAMVFSY	955
Db	894	GKIATEAIENPRTVVSLSLREQKFETMYAQSLOQIPYRNALKRAHVFITFAFTQAMVFSY	953
QY	956	ACCFRFGAYLVANEPNFDVLLVFSATVFGAMAVQVSSFADPYAKAKVSAAHVMIIIE	1015
Db	954	ACCFRFGAYLVARELMTFENVMVFSAVVFGAMAGNITSSFADPYAKAKVSASHIIGIE	1013
QY	1016	KSPIDSYSPHGLKPNTEGNTFENVFNPTRPDIPVLQGLSLEVKKGOTLALVGSSG	1075
Db	1014	KIPEIDSYSTGLKPNWLEGNVKNFNGVKNFNPTRPNIPVLQGLSFEVKGKOTLRLVGSSG	1073
QY	1076	CGKSTVOLLERFYDPLAGSLVDGKEIKHLNVQWMLRAHLGIVSQEPIFDCSAENIAY	1135
Db	1074	CGKSTVOLLERFYNPMAGTVFLDGKEIKQLNVQCVRA-LGIVSQEPIFDCSAENIAY	1132
QY	1136	GDNSRVVSHEEIMQAAKEANIHFITLPEKYNTRVGDGKGTQLSGGQKQRTAIARALVRQ	1195
Db	1133	GDNSRVVSHEEIVRAAREANIHFIDSLPEKYNTRVGDGKGTQLSGGQKQRTAIARALVRQ	1192
QY	1196	PHILLDEATSLDTESEKVVQVQALDKAREGRTCVIAHRLSTIQNADLIWVFQNGVKVE	1255
Db	1193	PHILLDEATSLDTESEKVVQVQALDKAREGRTCVIAHRLSTIQNADLIWVFQNGQVKE	1252
QY	1256	HGTHQOLLAQKIYFVSMVSVQAGAKR	1281
Db	1253	HGTHQOLLAQKIYFSM--VQAGAKR	1276
RESULT 7			
MDR3_HUMAN	STANDARD; PRT; 1279 AA.		
ID	MDR3_HUMAN		
AC	P21439;		
DT	01-MAY-1991 (Rel. 18, Created)		
DT	01-MAY-1991 (Rel. 18, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Multidrug resistance protein 3 (P-glycoprotein 3).		
GN	ABCB4 OR PGY3 OR MDR3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=89138016; PubMed=2906314;		
RA	van der Bliek A.M., Kooiman P.M., Schneider C., Borst P.;		
RT	"Sequence of mdr3 cDNA encoding a human P-glycoprotein.";		
RL	Gene 71:401-411(1988).		
RN	[2]		
RP	SEQUENCE OF 856-1279 FROM N.A.		
RA	MEDLINE=88111519; PubMed=2892668;		
RA	van der Bliek A.M., Baas F., ten Houte de Lange T., Kooiman P.M.,		
RA	van der Velde-Koerts T., Borst P.;		
RT	"The human mdr3 gene encodes a novel P-glycoprotein homologue and		
RT	gives rise to alternatively spliced mRNAs in liver.";		
RL	EMBO J. 6:3325-3331(1987).		

RN	[3]		
RP	GENE STRUCTURE.		
RX	MEDLINE=91161629; PubMed=2002063;		
RA	Lincke C.R., Smit J.J.M., van der Velde-Koerts T., Borst P.;		
RT	"Structure of the human MDR3 gene and physical mapping of the human		
RT	MDR locus.";		
RL	J. Biol. Chem. 266:5303-5310(1991).		
CC	-1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED		
CC	DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS. HUMAN MDR3 IS NOT		
CC	CAPABLE OF CONFERRING DRUG RESISTANCE. MEDIATES THE TRANSLLOCATION		
CC	OF PHOSPHATIDYLCHOLINE ACROSS THE CANALICULAR MEMBRANE OF THE		
CC	HEPATOCYTE.		
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.		
CC	-1- DISEASE: DEFECTS IN ABCB4 ARE A CAUSE OF PROGRESSIVE FAMILIAL		
CC	INTRAHEPATIC CHOLESTASIS TYPE III (PFIC), A FORM OF AUTOSOMAL		
CC	RECESSIVE LIVER DISORDERS, CHARACTERIZED BY EARLY ONSET OF		
CC	CHOLESTASIS THAT PROGRESSES TO CIRRHOSIS AND LIVER FAILURE BEFORE		
CC	ADULTHOOD.		
CC	-1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M23234; AAA36207.1; -.		
DR	EMBL; X06181; CAA29547.1; -.		
DR	PIR; JS0051; DVHU3.		
DR	PIR; A42213; A42213.		
DR	HSSP; P13359; INBD.		
DR	MIM; 171060; -.		
DR	MIM; 602347; -.		
DR	InterPro; IPR003593; AAA.		
DR	InterPro; IPR001140; ABC_transporter_tmem.		
DR	InterPro; IPR003439; ABC_transportr.		
DR	InterPro; IPR001687; ATP_GTP_A.		
DR	Pfam; PF00664; ABC_membrane; 2.		
DR	Pfam; PF00005; ABC_tran; 2.		
DR	SMART; SM00382; AAA; 2.		
DR	PROSITE; PS00211; ABC_TRANSPORTER; 2.		
KW	ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;		
KW	Multigene family.		
FT	DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).		
FT	TRANSMEM 58 78 POTENTIAL.		
FT	TRANSMEM 123 143 POTENTIAL.		
FT	TRANSMEM 192 211 POTENTIAL.		
FT	TRANSMEM 216 235 POTENTIAL.		
FT	TRANSMEM 301 320 POTENTIAL.		
FT	TRANSMEM 336 354 POTENTIAL.		
FT	DOMAIN 355 711 CYTOPLASMIC (POTENTIAL).		
FT	TRANSMEM 712 732 POTENTIAL.		
FT	TRANSMEM 756 776 POTENTIAL.		
FT	TRANSMEM 832 851 POTENTIAL.		
FT	TRANSMEM 854 873 POTENTIAL.		
FT	TRANSMEM 937 956 POTENTIAL.		
FT	TRANSMEM 976 993 POTENTIAL.		
FT	DOMAIN 994 1279 CYTOPLASMIC (POTENTIAL).		
FT	NP_BIND 429 436 ATP (BY SIMILARITY).		
FT	NP_BIND 1069 1076 ATP (BY SIMILARITY).		
FT	REPEAT 1 640		
FT	REPEAT 641 1279		
FT	CARBOHYD 91 91		
FT	CARBOHYD 97 97		
FT	CONFLICT 1093 1093		
SQ	SEQUENCE 1279 AA; 140682 MW; 3D58C98B5C8D6087 CRC64;		
Query Match 75.9%; Score 4915.5; DB 1; Length 1279;			
Best Local Similarity 75.7%; Pred. No. 3.2e-273;			
Matches 975; Conservative 131; Mismatches 163; Indels 19; Gaps 9;			


```
FT TRANSMEM 52 75 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 346 707 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 708 728 POTENTIAL.
FT TRANSMEM 753 773 POTENTIAL.
FT TRANSMEM 829 849 POTENTIAL.
FT TRANSMEM 850 870 POTENTIAL.
FT TRANSMEM 933 953 POTENTIAL.
FT TRANSMEM 970 990 POTENTIAL.
FT TRANSMEM 991 1276 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (BY SIMILARITY).
FT NP_BIND 1066 1073 ATP (BY SIMILARITY).
FT REPEAT 1 635
FT REPEAT 636 1276
SQ SEQUENCE 1276 AA; 140332 MW; A6C38DCD2C118EEF CRC64;

Query Match 75.1%; Score 4863; DB 1: Length 1276;
Best Local Similarity 74.8%; Pred. No. 3.2e-270;
Matches 960; Conservative 140; Mismatches 170; Indels 14; Gaps 8;

QY 1 MDPEGRKGSAAEK--NEWKMGKSKK--KEKKEKPTVTSTFAMFRYSNWLDRLYMLVGTMA 57
DB 1 MDLEARNGTARRLDGDFELGSIISNOGREKKKVNLIIGLLTLFRYSQDKLFWFLGTLT 60

QY 58 AIIHGAALPLMLVFGNWTDSFA--NAGISRNKTPFPVIINESINNTQHTNHLSEEMTY 116
DB 61 AIAHSGSLPLMLVFGNWTDSFA--NAGISRNKTPFPVIINESINNTQHTNHLSEEMTY 116

QY 117 AYYSGIGAGVLAAYIQVSWFCLAGQILKIRKQFFHAIHQEIGWDFVDVHVGELNTR 176
DB 114 AYYSGIGAGVLAAYIQVSWFCLAGQILKIRKQFFHAIHQEIGWDFVDVHVGELNTR 176

QY 177 LTDDVSKINGIGDKIGMFQSIATFTFGIVGTRGKLTVLAIASPVILGLSAIWA 236
DB 174 LTDDVSKINGIGDKIGMFQSIATFTFGIVGTRGKLTVLAIASPVILGLSAIWA 236

QY 237 ILSFTDKELLAYAKAGAAVEEVLAAIRTVIAFGGKKELERYKNKNLEAKGIGIKAIT 296
DB 234 ILSFTDKELLAYAKAGAAVEEVLAAIRTVIAFGGKKELERYKNKNLEAKGIGIKAIT 296

QY 297 ANISGAFLLIYASALAFWYGTSLVSEYSIGOVLTTFVFSVLIGAFSIGOASPSIEA 356
DB 294 ANISMGIAFLLIYASALAFWYGTSLVSEYSIGOVLTTFVFSVLIGAFSIGOASPSIEA 356

QY 357 FANARGAAYEIFKIIDNKNPSIDYSKSGHKPDNIKGNLFKNVHFSYPSRKEYKILKGLN 416
DB 354 FANARGAAYEIFKIIDNKNPSIDYSKSGHKPDNIKGNLFKNVHFSYPSRKEYKILKGLN 416

QY 417 LKVSQGTVALVNSCGKSTTVQLMQLRQYDPTDGMVCDIGQDIRTNVHRLREITGVVS 476
DB 414 LKVSQGTVALVNSCGKSTTVQLMQLRQYDPTDGMVCDIGQDIRTNVHRLREITGVVS 476

QY 477 QEPVLPAITTAENIRYGRNVTWDETEKAVKEANAYDFIMKPNKFTDLVGERGAOLSG 536
DB 474 QEPVLPAITTAENIRYGRNVTWDETEKAVKEANAYDFIMKPNKFTDLVGERGAOLSG 536

QY 537 OKORIAARALVRNPKILLDEATSAIDTSEAVVOVALDKARKGRTTIVIAHRLSTVRN 596
DB 534 OKORIAARALVRNPKILLDEATSAIDTSEAVVOVALDKARKGRTTIVIAHRLSTVRN 596

QY 597 ADVIAGFDGVIYKGNHDELMEKGIYKLVMTQTRGNEIELENATGSKSESDALEMS 656
DB 594 ADVIAGFDGVIYKGNHDELMEKGIYKLVMTQTRGNEIELENATGSKSESDALEMS 656

QY 657 PKDSGLIKRRSTRSIAHAPQODRKLTKEDLNENVPVSWRILKLNSTEWFPYVVG 716
DB 654 P-NGWKAIRFNSTRKSLKSPH--QNRIDETNELDANVPVSWRILKLNSTEWFPYVVG 716

QY 717 IFCAIINGGLQAPAFSIFRSIIGIFTRDEDPETKRONSNMFSVLFLVLGIISFIFFLQ 776
```

```
DB 712 TVCAIANGALQPAFSLIILSEMIAIFGPGDD-AVKQKQCNWFSVLGLGLVFFLFFLQ 770
QY 777 FTFGKAGEILTKRLRYMVRSMRLQDVSWFDDPKNTTGTALTRLANDAAQVKAIGSRJA 836
DB 771 FTFGKAGEILTKRLSRMAFKAMLRQDMSWFDHDKNSTGALSTRLATDAQAQVQATGKLA 830
QY 837 VITQNTANLGTGIIISLIYQWLTLLLAIVPIIAIAGVVMKMLSGQALKDKKELEGAG 896
DB 831 LIAQNTANLGTGIIISLIYQWLTLLLSVVPFIAVAGIVEMKMLAGNKRDKKEMEAG 890
QY 897 KIATEAIEENFTVSLTREQKFEYMYAQSLQVPRYNSLRKAHIFGVSFISITQAMVFSYA 956
DB 891 KIATEAIEENFTVSLTREQKFEYMYAQSLQVPRYNSLRKAHIFGVSFISITQAMVFSYA 950
QY 957 GCFRFGAYLVANEFMNFQDVLVFSYAIVFGAMAVGVSSFADYAKAKVSAABH7MIIEK 1016
DB 951 GCFRFGAYLVANEFMNFQDVLVFSYAIVFGAMAVGVSSFADYAKAKVSAABH7MIIEK 1016
QY 1017 SPLIDSYSPLGLKPNLTLEGNTFNEVFNYPTRDIPVLOGLSLEVKKGQTLALVSSGC 1076
DB 1011 QPLIDSYSPLGLKPNLTLEGNTFNEVFNYPTRDIPVLOGLSLEVKKGQTLALVSSGC 1070
QY 1077 GKSTVQVLLERFYDPLAGSLVDGKEIKHLNVQWLRHAGLIVSQBPILFDCSIAENIAYG 1136
DB 1071 GKSTVQVLLERFYDPLAGSLVDGKEIKHLNVQWLRHAGLIVSQBPILFDCSIAENIAYG 1130
QY 1137 DNSRVVSHHEIMOAAKEANIHHFIETPLPEKYNTRYVDGKTQSLSGGQKQRIATARALVRQ 1196
DB 1131 DNSRVVSHHEIMOAAKEANIHHFIETPLPEKYNTRYVDGKTQSLSGGQKQRIATARALVRQ 1190
QY 1197 HILLDEATSAIDTSEKVVQVQALDKARGRTCIVIAHRLSTIONADLIIVFPQNKVKEH 1256
DB 1191 HILLDEATSAIDTSEKVVQVQALDKARGRTCIVIAHRLSTIONADLIIVFPQNKVKEH 1250
QY 1257 GTHQQLLAQKGIYFMSVSVQAGAK 1280
DB 1251 GTHQQLLAQKGIYFMSVSVQAGAK 1274

RESULT 9
MDR3_CRIGR
ID MDR3_CRIGR STANDARD; PRT; 1281 AA.
AC P23174;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (P-glycoprotein 3).
GN PGY3 OR PGY3.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92135896; PubMed=1685679;
RA Endicott J.A., Sarangi F., Ling V.;
RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
   gene family."
RL DNA Seq. 2:89-101(1991).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC -!- DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT
CC CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
CC CANNOT.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
-----
CC EMBL; M60042; AAA68885.1; -.
CC DR HSSP; P13569; 1NBD.
CC DR InterPro; IPR003593; AAA.
CC DR InterPro; IPR001140; ABC.transprotet_tmern.
CC DR InterPro; IPR003439; ABC.transprotet.
CC DR InterPro; IPR001687; ATP.GTP_A.
CC DR Pfam; PF00664; ABC_membrane; 2.
CC DR Pfam; PF00005; ABC_tran; 2.
CC DR SMART; SM00382; AAA; 2.
CC DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
CC KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 299 319 POTENTIAL.
FT TRANSMEM 328 348 POTENTIAL.
FT DOMAIN 349 712 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 713 733 POTENTIAL.
FT TRANSMEM 758 778 POTENTIAL.
FT TRANSMEM 834 854 POTENTIAL.
FT TRANSMEM 855 875 POTENTIAL.
FT TRANSMEM 938 958 POTENTIAL.
FT TRANSMEM 975 995 POTENTIAL.
FT DOMAIN 996 1281 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 429 436 ATP (POTENTIAL).
FT NP_BIND 1071 1078 ATP (POTENTIAL).
FT SEQUENCE 1281 AA; 140866 MW; 2203EF61EB29602 CRC64;

Query Match 74.5%; Score 4822.5; DB 1; Length 1281;
Best Local Similarity 73.7%; Pred. No. 6.6e-268;
Matches 949; Conservative 148; Mismatches 174; Indels 17; Gaps 8;

QY 1 MDPEGKKGSAEK-----NFWKKGKSKK-KEKKKKPTVSTFAMFRYSNWLDRMLVVG 54
DB 1 MDLEAARNGTARRPTGVGEFGLGSLNNGRNKKKKVNLGLPLTFRYSWDQKFLMLLG 60

QY 55 TMAAIIHGAALPLMLVFGNMTDSFA-NAGISRNKTFPVIINESINTNTQHFINHLSEEM 113
DB 1 TMAIAHGSGLPLMLVFGEMTDKFEVNNAG---NFSLPVNFSLSMINPGR----ILEEEM 113

QY 114 TTYAYTSGAGVVAAYIQVSWCLAGROTLKTRKOFFHAIMRQELGWDFVDVHVGEL 173
DB 114 TRYAYTSGGLGVVAAYIQVSWCLAGROTLKTRKOFFHAIMRQELGWDFVDVHVGEL 173

QY 174 NTRLTDDVSKINGIGDKIGMFOQSIATFTGTFVGTGRGWLTLVILAIISPVLGLSAAI 233
DB 174 NTRLTDDISKISGIGDKVGMFFQAVATFAGFVIGRWKLTVMIAISPILGLSAV 233

QY 234 WAKILSFTDKELLAYAKAGAAVEELAAIRTVIAFGQKKKELERYKNKLEAKGIGIKK 293
DB 234 WAKILTSFDSKELAAAYAKAGAAVEELAAIRTVIAFGQKKKELERYQKHLNKAAGIKK 293

QY 294 AITANISGAFLIIYASYALAFWYGTSLVLSSEYSIGQVLTFFVFSVLGAFSIGOASP 353
DB 294 AITANISMGTAFLIIYASYALAFWYGTSLVLSSEYSIGQVLTFFVFSVLGAFSIGOAPC 353

QY 354 IEAFANARGAAEYIFKIDNKPISDYSKSGHKPDNKNLEFKNVHFSPYSRKEVKILK 413
DB 354 IDAFANARGAAEYIFDIIDNKPISDYSKSGHKPDNKNLEFSDVHFSPYSRANIKILK 413

QY 414 GLNLKVGSGTVALVNGSGCGKSTTVOLMORLYDPTDGMVCDIGQDRTINVRHLREITG 473
DB 414 GLNLKVGSGTVALVNGSGCGKSTTVOLMORLYDPTDGMVCDIGQDRTINVRHYLEITG 473

QY 474 VVSGPEVLFATTAENIRYGRNVTWDEIEKAVKEANAYDFIMKLNKPDFTLVGERGAOL 533
DB 474 VVSGPEVLFATTAENIRYGRNVTWDEIEKAVKEANAYDFIMKLNKPDFTLVGERGAOL 533
```

```
DB 474 VVSGPEVLFSTTTAENIRYGRNVTWDEIEKAVKEANAYDFIMKLNKPDFTLVGERGAOL 533
QY 534 SGGQKQRIATARALVRNPKILLDDATSEALDSEAVVQVALDKARKGRITIVIAHRLST 593
DB 534 SGGQKQRIATARALVRNPKILLDDATSEALDSEAEVQALDKAREGRITIVIAHRLST 593
QY 594 VRNADVIAGDDGVIVKGNHDELMKEGKYFKLVNMTQTRNGNEIELENATGESKSDAL 653
DB 594 VRNADVIAGDDGVIVKGNHDELMKEGKYFKLVNMTQTRNGNEIELENATGESKSDAL 653
QY 654 EMSPKDSGSSLIKRRSTRSRTHAPOGODRKLGTEK-LNENVPVPSFWRLKLNSTWEPY 712
DB 654 GMP-NGWKSHIFRNSTKSLKSRRAHHRLDVEDDDELDAENVPVPSFWRLKLNSTWEPY 712
QY 713 FVVGIFCAITNGGLQPAFSIIFRSIIIGFTIRDEDEPTKRONSMFVFLVGLGIISFITF 772
DB 713 FVVGIVCAIVNGALQPAISIIISEMIAIFGPGDD-AVKQCKNLFSVLVGLGLVLSFFTE 771
QY 773 FLOGFTFGKAGEILTKRLRYMFRSMRLRQDSVDFDDPKNTGALTTRLANDAAQVKGAG 832
DB 773 FLOGFTFGKAGEILTKRLRYMFRSMRLRQDSVDFDDPKNTGALTTRLANDAAQVKGAG 831
QY 833 SRLAVITQNTANLGTIIISLYGWQLTLLLAIVPIIATAGVVMKMLSGQALKDKKEL 892
DB 833 TRLALIAQNTANLGTIIISLYGWQLTLLLAIVPIIATAGVVMKMLSGQALKDKKEL 891
QY 893 EGAGKIATEATENRTVVSLSLTREQFEYMAQSLQVYPYNSLRKAHIFGVFSFISQAMMY 952
DB 893 EAAGKIATEATENRTVVSLSLTREQFEYMAQSLQVYPYNSLRKAHIFGVFSFISQAMMY 951
QY 953 FSYACGRFYGAYLVANFEMFQDVLVLSAIVFCAMAVGVSSFPADYAKAKVSAAHVIM 1012
DB 953 FSYACGRFYGAYLVANFEMFQDVLVLSAIVFCAMAVGVSSFPADYAKAKVSAAHVIM 1011
QY 1013 IIEKSLDLSYSPHGLKPTLEGNVTFENVFNTPRDPVQLQGLSLEVKKGQTLALVG 1072
DB 1013 IIEKSLDLSYSPHGLKPTLEGNVTFENVFNTPRDPVQLQGLSLEVKKGQTLALVG 1071
QY 1073 SSGCGKSTVQVQLLRFYDPLAGSLVDIGKEIKHLNWLRAHILGIVSQEPIILFDCSTAE 1132
DB 1073 SSGCGKSTVQVQLLRFYDPLAGSLVDIGKEIKHLNWLRAHILGIVSQEPIILFDCSTAE 1131
QY 1133 IAYGDSNVVSHSEETMOAAKEANLHRTETLPEKYNTRVGDGKGTOLSGGQKQRIATARAL 1192
DB 1133 IAYGDSNVVSHSEETMOAAKEANLHRTETLPEKYNTRVGDGKGTOLSGGQKQRIATARAL 1191
QY 1193 VRQPHILLDEATSEKVVQEAIDKAREGRTICIVIAHRLSTIONADLIIVVFONGK 1252
DB 1193 VRQPHILLDEATSEKVVQEAIDKAREGRTICIVIAHRLSTIONADLIIVVFONGK 1251
QY 1253 VKEGTHOOLLAQKGIYFSMVSVQAGAK 1280
DB 1253 VKEGTHOOLLAQKGIYFSMVSVQAGAK 1279

RESULT 10
MDR2_RAT STANDARD; PRT; 1278 AA.
AC Q08201;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 2 (P-glycoprotein 2).
GN PGY2 OR MDR2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISSCHER;
RX MEDLINE=93376516; PubMed=8103593;
RA Brown P.C., Thorgerisson S.S., Silverman J.A.;
```

"Cloning and regulation of the rat mdr2 gene.";
Nucleic Acids Res. 21:3885-3891(1993).
-!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L15079; AAA02937.1; .
HSSP; P13569; INED.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane_2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
Multigene family.
FT DOMAIN 1 54 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 55 75 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT DOMAIN 346 709 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 710 730 POTENTIAL.
FT TRANSMEM 755 775 POTENTIAL.
FT TRANSMEM 831 851 POTENTIAL.
FT TRANSMEM 852 872 POTENTIAL.
FT TRANSMEM 935 955 POTENTIAL.
FT TRANSMEM 972 992 POTENTIAL.
FT DOMAIN 993 1278 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (POTENTIAL).
FT NP_BIND 1068 1075 ATP (POTENTIAL).
SQ SEQUENCE 1278 AA; 140655 MW; DAFB6BE745AF73BF CRC64;

Query Match 74.3%; Score 4809; DB 1; Length 1278;
Best Local Similarity 73.8%; Pred. No. 3.9e-267;
Matches 948; Conservative 147; Mismatches 176; Indels 14; Gaps 8;
QY 1 MDPEGRKGSAAK--NEWKWK-KSKKKKKKKKPTVSTFAMPRYSNWLDRLMLVGTMA 57
Db 1 MDLEAARNGTARRLDGDFELGSGTSNQREKKKYNLGLPLTFRYSWQDKLFMLGTAM 60
QY 58 AIIHGAALPLMLVGNMTDSFA-NAGISRNKTPFPVPIINESITNNTQHFHNLHEEMTY 116
Db 61 AIAHSGGLPLMIVFGEMTDKFDVNDAG---NFSLPVNFSLMNPGR----ILEEMTRY 113
QY 117 AYYTSGIGAGVLVAAYTOVSFWCIAAGRQILKIRKQFFHAIMRQEGISGFVHDVGNLNR 176
Db 114 AYYTSGIGAGVLVAAYTOVSFWCIAAGRQILKIRKQFFHAIMRQEGISGFVHDVGNLNR 173
QY 177 LTDDVSKINEGIGDKIGMFTQSTATFTTGTIVGTFTRGWKLTIVLAISPVLGSAIAWAK 236
Db 174 LTDDISKISEGIGDKVGMFFQATFTFAGTIVGTFIRGWKLTIVLAISPVLGSAIAWAK 233
QY 237 ILSSTOKELLAYAKAGAAVEVLAAITRTVAFGGKKELEKYNKNEAKGIGKAIT 296
Db 234 ILSFTSKELAAAYAKAGAAVEALGALRTVAFGGKKELEKYNKNEAKGIGKAIT 293
QY 297 ANISGAFLIYASYALAFWYGTSLVLSSEYSIGQVLTFFVSFVIGAFSIGQASPSIEA 356
Db 294 ANISMGIAFLIYASYALAFWYGTSLVLSSEYSIGQVLTFFVSFVIGAFSIGQASPSIEA 353

QY 357 FANARGAAYELFKLIDNKPISIDYSKSGHKPDNKGKLEPKNVHFSYPSRKEVKILKGLN 416
Db 354 FPNARGAAYVIFIDNNPKIDSFSERGHKPDISKGNLEFSDFVHFSYPSRANIKILKGLN 413
QY 417 LKVSOGTVALVNGSGCGKTTVQLMORLYDPTDGMVCDIGDIRTNVHRLREITGVVYS 476
Db 414 LKVSOGTVALVNGSGCGKTTVQLMORLYDPTDGMVCDIGDIRTNVHRLREITGVVYS 473
QY 477 QEPVLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKEDTTLVGERGAQLSG 536
Db 474 QEPVLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKEDTTLVGERGAQLSG 533
QY 537 QKQRIATARALVRNPKILLDEATSAIDSEAVQVVALDKARKGTTIVIAHRLSTVRN 596
Db 534 QKQRIATARALVRNPKILLDEATSAIDSEAVQVVALDKARKGTTIVIAHRLSTVRN 593
QY 597 ADVIAGFDDGVIIVEKGNHDELMKEGIYFKLVTMQRGNEIELENATGESKESDALEMS 656
Db 594 ADVIAGFDDGVIIVEKGNHDELMKEGIYFKLVTMQRGNEIELENATGESKESDALEMS 653
QY 657 PKDSGSLIKRRTRRSIHAPOQDQRLGKTK-EDLNENVPVPSFWRILKLNSTENWPFYV 715
Db 654 P-NGWKARIFRNTKSLKSSRAHQNRDLVETNELDANVPVPSFWRILKLNSTENWPFYV 712
QY 716 GIFCAIINGLOPAFSLIIFSRIGIETREDDETPETKRONSMFSLVFLVGLIITFFLQ 775
Db 713 GTLCAIANGALOPAFSLIIFSRIGIETREDDETPETKRONSMFSLVFLVGLIITFFLQ 771
QY 776 GFTFGKAGEILTTLRLYMYFRSMRDVSNFDDPKNTGALTTLRLANDAQAQVKGAGSL 835
Db 772 GFTFGKAGEILTTLRLYMYFRSMRDVSNFDDPKNTGALTTLRLANDAQAQVKGAGSL 831
QY 836 AVITQNIANGTGIISLIYQWLTLLLAIVPITIAIGVVMKMLSGQALKDKKELEGA 895
Db 832 ALIAQNTANLGTGIISLIYQWLTLLLAIVPITIAIGVVMKMLSGQALKDKKELEGA 891
QY 896 GKATATAIENIRTVSILTRQEFYMYAQSLOVPYNSLRKAHIFCVSFTSIQAMMYFSY 955
Db 892 GKATATAIENIRTVSILTRQEFYMYAQSLOVPYNSLRKAHIFCVSFTSIQAMMYFSY 951
QY 956 AGCFRGAYLVANEFMNFODVLLVFSIAIVFGAMAVGOVSSEAPDYAKAKVSAHVIMITE 1015
Db 952 AGCFRGAYLVANEFMNFODVLLVFSIAIVFGAMAVGOVSSEAPDYAKAKVSAHVIMITE 1011
QY 1016 KSPILDSYSPHGLKPNLTGEGNVTFNEVFNYPTRPDIPQLGLSLEVKKQTLALVSGSG 1075
Db 1012 RQPLDSYSRSGMWPDKFEGSVTFNEVFNYPTRPDIPQLGLSLEVKKQTLALVSGSG 1071
QY 1076 CGKSTVVQLLERYDPLAGSLVDGKEIKHLNVQWRAHLGIVSQBPILFDCSIAENIAY 1135
Db 1072 CGKSTVVQLLERYDPLAGSLVDGKEIKHLNVQWRAHLGIVSQBPILFDCSIAENIAY 1131
QY 1136 GDNRSVVSHEEIMQAAKEANIHFFIETLPEKYNTRYDGTOLSGGOKORIAIALRVQ 1195
Db 1132 GDNRSVVSHEEIMQAAKEANIHFFIETLPEKYNTRYDGTOLSGGOKORIAIALRVQ 1191
QY 1196 PHILLDEATSAIDSEKVVQVQALDKAREGRTCTIVIAHRLSTIQNADLIIVFQNGKVE 1255
Db 1192 PHILLDEATSAIDSEKVVQVQALDKAREGRTCTIVIAHRLSTIQNADLIIVFQNGKVE 1251
QY 1256 HGTHQQLLAQKGIYFSWVSQAGAK 1280
Db 1252 HGTHQQLLAQKGIYFSWVSQAGAK 1276
RESULT 11
AB11_HUMAN
ID AB11_HUMAN STANDARD; PRT; 1321 AA.
AC O95342; Q9UNB2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11).

GN ABCB11 OR BSEP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A., AND VARIANT PFC2 GLY-297.

RX MEDLINE=99021377; PubMed=9806540;

RA Strauthle S.S., Bull L.N., Knisely A.S., Kocoshis S.A., Dahl N.,

RA Arnell H., Sokal E., Dahan K., Childs S., Ling V., Tanner M.S.,

RA Kagaliwalla A.F., Nemeth R.M., Pawlowska J., Baker A., Miell-Vergani G.,

RA Freilmer N.B., Gardiner R.M., Thompson R.J.;

RT "A gene encoding a liver-specific ABC transporter is mutated in

RL progressive familial intrahepatic cholestasis.";

RL Nat. Genet. 20:233-238(1998).

RN [2]

RP SEQUENCE FROM N.A.

RA Mol O., Hooiveld G.J.E.J., Jansen P.L.M., Muller M.;

RA "Cellular localization and functional characterization of the human

RT bile salt export pump (BSEP).";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS

CC INTO THE CANALICULUS OF HEPATOCYTES.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN

CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR

CC MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN

CC SITU (BY SIMILARITY).

CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,

CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN

CC ATP BINDING CASSETTE (ABC) DOMAIN.

CC -1- DISEASE: DEFECTS IN ABCB11 ARE THE CAUSE OF PROGRESSIVE FAMILIAL

CC INTRAHEPATIC CHOLESTASIS 2 (PFC2), AN INHERITED LIVER DISEASE OF

CC CHILDHOOD. PFC2 IS CHARACTERIZED BY CHOLESTASIS AND NORMAL SERUM

CC GAMMA-GLUTAMYLTRANSFERASE ACTIVITY.

CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF091582; AAC77455.1; -

DR EMBL; AF136523; AAD28285.1; -

DR MIM; 603201; -

DR MIM; 601847; -

DR InterPro; IPR003593; AAA.

DR InterPro; IPR001140; ABC_transporter_tmnm.

DR InterPro; IPR003439; ABC_transporter.

DR InterPro; IPR001687; ATP_GTP_A.

DR Pfam; PF00664; ABC_membrane; 2.

DR Pfam; PF00005; ABC_tran; 2.

DR SMART; SM00382; AAA; 2.

DR PROSITE; PS00211; ABC_TRANSPORTER; 1.

KW ATP-binding; Transmembrane; Transport; Disease mutation.

FT DOMAIN 1 62 CYTOPLASMIC (POTENTIAL).

FT TRANSEM 63 83 POTENTIAL.

FT DOMAIN 84 147 EXTRACELLULAR (POTENTIAL).

FT TRANSEM 148 168 POTENTIAL.

FT DOMAIN 169 215 CYTOPLASMIC (POTENTIAL).

FT TRANSEM 216 236 POTENTIAL.

FT DOMAIN 237 240 EXTRACELLULAR (POTENTIAL).

FT TRANSEM 241 261 POTENTIAL.

FT DOMAIN 262 319 CYTOPLASMIC (POTENTIAL).

FT TRANSEM 320 340 POTENTIAL.

FT DOMAIN 341 353 EXTRACELLULAR (POTENTIAL).

FT TRANSEM 354 374 POTENTIAL.

FT DOMAIN 375 755 CYTOPLASMIC (POTENTIAL).

FT TRANSEM 756 776 POTENTIAL.

FT	DOMAIN	777	794	EXTRACELLULAR (POTENTIAL).
FT	TRANSEM	795	815	POTENTIAL.
FT	DOMAIN	816	869	CYTOPLASMIC (POTENTIAL).
FT	TRANSEM	870	890	POTENTIAL.
FT	TRANSEM	891	911	POTENTIAL.
FT	DOMAIN	912	979	CYTOPLASMIC (POTENTIAL).
FT	TRANSEM	980	1000	POTENTIAL.
FT	DOMAIN	1001	1011	EXTRACELLULAR (POTENTIAL).
FT	TRANSEM	1012	1032	POTENTIAL.
FT	DOMAIN	1033	1321	CYTOPLASMIC (POTENTIAL).
FT	NP_BIND	455	462	ATP (POTENTIAL).
FT	NP_BIND	1113	1120	ATP (POTENTIAL).
FT	CARBOHYD	109	109	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	122	122	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	297	297	E -> G (IN PFC2).
FT				/FtId=VAR_010271.
FT	CONFLICT	339	339	V -> L (IN REF. 2).
FT	CONFLICT	444	444	V -> A (IN REF. 2).
SEQ	SEQUENCE	1321 AA;	146392 MW;	D444CACC48DEA371 CRC64;

Query Match 51.0%; Score 3302.5; DB 1; Length 1321;

Best Local Similarity 50.7%; Pred. No. 5.2e-181;

Matches 653; Conservative 242; Mismatches 359; Indels 33; Gaps 9;

QY	20	KKSK-KKEKKKPTVSTFAMFRYSNWLDRYMLVGMMAIIHGAALPLMLVFGNMTDS	78
DB	30	KKSLQDEKKGQVGRVGFQLFREFSSSDIWLFMFVGSICAFGLTAQPGVLLIFGTMTDV	89
QY	79	FANAGISR-----NKTTF---PVIINESLTNTQHEINHLREEMTYAYVYSGIG	124
DB	90	FDYDVELOEIQIPGKACVNNTIWNSSLNQNMNTGTRCSLLWIESEMIKFASYAGIA	149
QY	125	AGVLVAAYIQSVFWCLAAAGRIKIRKQFHAIIMRQIEGDFVDHVGELNRLTDDYSKI	184
DB	150	VAVLITGYIQCWFVIAAARQIQMKRKYFRIMRMEIGWDFDCNSVSELNTRFSDINKI	209
QY	185	NEGIDKTMGFQSTATFTFTGVTGKWLTVLAISVPLGSLAAIAWAKILSSFTDK	244
DB	210	NDAIQDALFTQRTSTICGFLGFRGKLTIVTSPLIGIAGATIGLSVSKFTDY	269
QY	245	ELLAYAKAGAAVEELAAIRTVIAFGGOKKELERYKNLEAKGIGIKKAITANISIGAA	304
DB	270	ELKATAGVVADEVISSMRVVAAGGKREVEREKNLVFAQRWIRKGVIMGFFTGTV	329
QY	305	FLLIYAYALAFWYGTSLVL--SSEYSIGQVLTVPFVSLVIGAPSGQASPSFEAFANARGA	363
DB	330	WCLIFLCYAVAFWYGSTLVLDGEYTPGTQLVQIFLSVIVGALNLGNASPCLEAFATGRAA	389
QY	364	AYEIFKIDNKPISDSYKSGHKPNIKGNLEFKNVHESYPSRKEVTKLGLNLKVOSSQ	423
DB	390	ATSFETIDRKPIDDCMEDGYKLDRIKGEIEFHNVTHYPSRPEVKILNDLNVIRKPE	449
QY	424	TVALVNSGCGKSTTVOLMQRLYDPTDGMVCIDGDQDIRTINVRLHREITGVVSEPVLFA	483
DB	450	MTALVPGSGAGKSTALQIQRFYDPCGMYVDGHDHISLNLQWLQDQIGIVEQEPVLS	509
QY	484	TTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGREGAQLSGGQKQRTAI	543
DB	510	TTIAENIRYGRDATMEDIVQAAKEANAYNFIMDLPPQOFDTLVGEGGQMGSGQKQRTAI	569
QY	544	ARALVNRPKILLDDATSAIDTESAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGF	603
DB	570	ARALVNRPKILLDDATSAIDNESAMVQEVLSKIQHGTITISVAHRLSTVRAADTIIGF	629
QY	604	DDGVIVEKGNHDELMKEGIVFKLVTMOTRG----NETELENAETGESKESDALEMSPKD	659
DB	630	EHGTAVERGTHEELLERKGVFTLVTLQSGNQNALNEEDIKDADDDMLARTFSRGSYQD	689
QY	660	SGSSLIKRRSTR---SIHAP--QGQDRKLGTKE-----LNENVPVPSFWRLTKLNS	707
DB	690	SLRASIRQRSKLSQSLVYLVEHPPLAVVDVHKSTVEEDKDKDIPVQEVPEPAPVRRILKFS	749

```
QY 708 TEMPVYVGVIFCAIINGLQPAFSAIIFSIIGIIFTRQEDPDKTRONSMFSLVFLVGLII 767
DQ 750 PEMVYMLVSGVAANGVITPLAFLESIQILGTFSPDKREQSQINGVCLLVFVANGCV 808
QY 768 SFITFFLOGFTFGAGEILTKRLRYMVRSMRDVDFDPKNTGALTTRLANDAAQV 827
DQ 809 SLFTQFLQYAFKSGELLTKRLKRFGRFAMLGQDIAWFDDLRNPSGALTTRLATDASQ 868
QY 828 KGATGSLRAVITONTANLGTGIIISLYGWQTLTLLLAIVPIIAIAGVEMKMLSGQALK 887
DQ 869 QGAGSQIGMIVNFTNTVAMIAIFASFWSKLSLVILCFEFTALSGATQTRMLTGFSR 928
QY 888 DKLEAGAGKIATAEIENFRVWSLITREKQFEYMAQSLQVPRYNSLRSLKHAHIFGVSPSIT 947
DQ 929 DKQALEMVGQITNEALSNIIRVAGIGKERRFIEALETELEKPPFTATOKANIYGFCAFA 988
QY 948 QAMVFSYAGCFRGAYLVANEFNFOVDLVLSAIVFGAMAVQGVSSFAPDYAKAKVSA 1007
DQ 989 QCIMFIANSASYRGYLLISNEGLHFSYVTRFVISAVALSATALGRAFSYTPSYAKAKISA 1048
QY 1008 AHVIMILEKSLIDYSPHGLKPNLTGNTVFNEVFNYPTRDIPVLOGLSLEVKKQOT 1067
DQ 1049 ARFQLIDRQPIISVNTAGEKWNDFQKIDFVCKFTYPSRPSQVLSVSPGOT 1108
QY 1068 LALVSSGCGKSTVVQLLRYFDPVLAGSLVDGKEIKHLNVQWLRHLGIIVSOEPIFLDC 1127
DQ 1109 LAFVSSGCGKSTSIQLLRYFDPVQKVMIDGSHDKVNVQFLRNSIGIVSEPEVLFAC 1168
QY 1128 SIANIAGDNRVSVSHEEIMQAANEIHHFIEITLPEKYNTRYGDKGTOLSGGQKORIA 1187
DQ 1169 SIMNIRYGDNTKEIPMERVIAAAQQLHDFVMSLPEKYETNVSGSQSLRGEKORIA 1228
QY 1188 IARALVROPHILLDEATSAIDTESKVVQEAIDKAREGTCIVIAHRLSTIQNADLIVV 1247
DQ 1229 IARRAIVRDKILLDEATSAIDTESKTVQALDKAREGTCIVIAHRLSTIQNADIIAV 1288
QY 1248 FQNGKVEHGHQQLLAQKGIYFSMVS 1274
DQ 1289 MAQGVVIEKGHEELMAQKAYYKLVIT 1315

RESULT 12
AB11_RABIT STANDARD; PRT; 1321 AA.
AC Q9NOV3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11)
DE (Sister of p-glycoprotein).
GN ABCB11 OR BSEP OR SPGP.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
RA Balasubramanian N.V., Suchy F.J., Ananthanarayanan M.;
RT "Molecular cloning and characterization of rabbit liver bile salt
RT export pump (Bsep/spgp).";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
CC -1- INTO THE CANALICULUS OF HEPATOCYTES.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
CC MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
CC SITU.
CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
```

```
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF249879; AAF65552.1; -.
CC InterPro: IPR003593; AAA.
CC InterPro: IPR001140; ABC_transporter_tmem.
CC InterPro: IPR003439; ABC_transportr.
CC InterPro: IPR001687; ATP_GTP_A.
CC Pfam: PF00864; ABC_membrane; 2.
CC Pfam: PF00005; ABC_tran; 2.
CC SMART: SM00382; AAA; 2.
CC PROSITE: PS00211; ABC_TRANSPORTER; 1.
CC ATP-binding; Transmembrane; Transport;
CC DOMAIN 1 62 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 63 83 POTENTIAL.
CC DOMAIN 84 147 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 148 168 POTENTIAL.
CC DOMAIN 169 215 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 216 236 POTENTIAL.
CC DOMAIN 237 240 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 241 261 POTENTIAL.
CC DOMAIN 262 319 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 320 340 POTENTIAL.
CC DOMAIN 341 353 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 354 374 POTENTIAL.
CC DOMAIN 375 755 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 756 776 POTENTIAL.
CC DOMAIN 777 794 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 795 815 POTENTIAL.
CC DOMAIN 816 869 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 870 890 POTENTIAL.
CC TRANSMEM 891 911 POTENTIAL.
CC DOMAIN 912 979 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 980 1000 POTENTIAL.
CC DOMAIN 1001 1011 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 1012 1032 POTENTIAL.
CC DOMAIN 1033 1321 CYTOPLASMIC (POTENTIAL).
CC NP_BIND 455 462 ATP (POTENTIAL).
CC NP_BIND 1113 1120 ATP (POTENTIAL).
CC CARBOHYD 109 109 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 122 122 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 1321 AA; 146376 MW; 457539 FCD6D717A2 CRC64;
CC Query Match 49.9%; Score 3230.5; DB 1; Length 1321;
CC Best Local Similarity 49.5%; Pred. No. 6.8e-177;
CC Matches 639; Conservative 249; Mismatches 362; Indels 41; Gaps 8;
QY 20 KKSKEKKEKKEKPT-VSTFAMERYSNWLDRLYMLVCTMAAIIHGAALPLMLVFGNWDTS 78
DQ 30 KKSRLQDKKKSQSVRFQFLFRFSWTDIWMCMGSLCACIHGIAQGVLLIFGTMTDV 89
QY 79 FANAGISR-----NKTFFVI---INESITNTQHFNFHLEEMTVVAYYSIG 124
DQ 90 FIDYDELQELKIPGKACVNNIVNSSLNQNVNNGTCGLGLDIESEMRFAGYAGIG 149
QY 125 AGVLVAAYIQVSWCLAGROILKIRKQFFHAIMRQETGWDFVDVGBELTFLTDVSKI 184
DQ 150 IAVLTGTGTCICFWGIAAAHQIKMKSYFRKIMRMGIGVDCNSVGKLNTPFSVDENKI 209
QY 185 NESIGDKIGMFPQSTATFTTGFIVGFTGRWKLTVILAIKSPVGLSAIAWAKILSSFTDK 244
DQ 210 NDSSADQLAIFIQGTMTSPFGLVGSQWKLTVIISPLIGLGAAILGLSVSKFTDY 269
QY 245 ELLAYAKAGAAEVLAAIRTVIAFGQKKELERYKNKLEAKGIGIKKAITANISGAA 304
```



```
Db 270 ELKAYAGSVADEISSNRTVAATGGEKKEVERVEKNLVFAQRGIRKGIWMGFETGYM 329
QY 305 FLIIYASYALAFWYGSIVL-SSEYSIGQVLTVFFSVLIGAFSICQASPSIEAFANARGA 363
Db 330 WCLIFCYALAFWYGSIVL-EEGYSIPGALVQIFLSVIGALNLGNASPCLEAFAAGRA 389
QY 364 AYEIKIINDKSIDSYSGSHKPNKGNLEFKNFHVSYSRKEVKILGLNLQVSGQ 423
Db 390 ASSIFETIDRKPIIDCMSEDGKLEKRIKEIEFHNWTHYPSRPEVKILNLSWIKYGE 449
QY 424 TVALVNSGCGSTTVQLMORLYDPTDGMVCIDGDIRTYNVRHLREITGVVYSQBPVIFA 483
Db 450 MTAIVGSPGAGKSTALQIHRFYPTGPMVTVESHDIRSHIQWLNRNQIGIVEQBPVLFF 509
QY 484 TTAENIRVGNVNTWDETEKAVKEANAYDFIMKLPNKEDTLVGERGALSGGQKORTAI 543
Db 510 HTIAEKIRYGREDAWEDTIOAKEANAYNFINDLPQQDFTLVGEQGMGSGQKORVAI 569
QY 544 ARALVRNPKILLDEATSDALDSEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGF 603
Db 570 ARALIRNPKILLDMATSDALDSEAVQVALSKTOHGHTIVSVAHRPATIRTDVILGC 629
QY 604 DGVVIVEKGNHDELMKEGIIYKLVMTQTRGNEIELE----NATGESKSESDALEMSPKD 659
Db 630 EHGAVERGTTEELLERKGVYFALVTLQSRNQGDQEEKEDATEDIDPEKTFSRGNQD 689
QY 660 SGGSLKRRS-----TRSHIAPQODRKLTKEDNLNVPVPSFWRIL 703
Db 690 SLRASLRQKRSQSLSYLAHEPPMAVEDHKSTHEEDRKDLPAQED----IEPASVRRIM 745
QY 704 KLNSTWEPVFWGIFCAIINGLQPAFSIIFRSIIGIFTRDEDPETKRONSMFSLFIV 763
Db 746 KLNAPWPMYMLSGMAVNGAVTPLYAFLFSQILGTFSL-PDKEQRSGQINGICLLFVT 804
QY 764 LGIISFTFFLQGTFTGKAGEIILTKRLRYWVFMSLRQDVWDFDPKNTTGALTITRLAND 823
Db 805 LGCVSFTQPLQGTFTAKSGELLTKRLKFGFRAMLGDIGWDFDLRNSFGALTITRLAID 864
QY 824 AAOVKGAIGSLRAVITONTANLGTGIIISLIYGMQVTLALLAIVPIATAGVYEMKMLSG 883
Db 865 ASOQVQATGSGIQGMWNSFTNVTVAMIAFLFSWKUTLGVICFFPPLALSGALQIKMLTG 924
QY 884 QALKDKKELSGAGKIATEIENRTVYSLTREQKFYMYAQSLQVYRNSLRKAHIFGVS 943
Db 925 FASDRQALEKAGQITSEALSNIRTVAGIKGERKFTETPEALEKPYKMAIKANVYGLC 984
QY 944 FSTQAMWYSYAGCFRFGAYIVANEFMNFODVLLVFSALVFGAMAGQVSSFAPDYAKA 1003
Db 985 FGFSQCITFTIANSASYRYGGYLLISNEGLHFYSYVFRVISAVVLASATALGRASSYTPSYAKA 1044
QY 1004 KVSAAHVIMIEKSPIDSYSPHGLKPNTLEGWNTFENVVFNYPTRPDIPVLOGLSLEYK 1063
Db 1045 KISAARFQLLDKQPPINLVYSSAGEKWDNFGKIDFVCKFTYPSRDPDLOVLNGLSVMS 1104
QY 1064 KGOTLALVSGSGCKSTVQVLLERFYDPLAGSVLIDGKEIKHLNVQWLAHGLGIVSQEPI 1123
Db 1105 PROTLAFVSGSGCKSTSIOLLERFYDPDHGKVMIDGHDHSKRVNIQFLRNSIGIVSQEPI 1164
QY 1124 LFDCSIAENTAYGNSRVWSHEEIMQAQKANTHHETLPEKYNTFRVGDGTQLSGGQK 1183
Db 1165 LFACSIKDNKYGDNTQETPEIRIIAAAKAQAQVHDFVMSLPEKYETNVGSGQSLSRGEK 1224
QY 1184 QRTAIALRVROPHILLDEATSDALDSEKVVQVQALDKAREGRTCVIAHRLSTIQNAD 1243
Db 1225 QRTAIAARIVRDPKILLDDEATSDALDSEKTVQVQALDKAREGRTCVIAHRLSTIQNSD 1284
QY 1244 LIVVFONGKVEHGTHTQOLLAQKGIYFSMVS 1274
Db 1285 IIAVMSQGMVTEKGTBEELAVQKGYKLV 1315
```

RESULT 13

```
AB11_MOUSE
ID AB11_MOUSE STANDARD; PRT; 1321 AA.
AC Q9QY30; Q9QZE8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11)
DE (sister of P-glycoprotein).
GN ABCB11 OR BSEP OR SPGP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=20076398; PubMed=10607905;
RA Green R.M., Hoda F., Ward K.L.;
RT "Molecular cloning and characterization of the murine bile salt export
pump.";
RL Gene 241:117-123(2000).
[2]
RP SEQUENCE OF 463-635 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RA Salkar R., Suchy F.J., Ananthanarayanan M.;
RT "Molecular cloning of mouse liver bile salt export pump (bsep).";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
INTO THE CANALICULUS OF HEPATOCYTES.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
SITU.
CC -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
ATP BINDING CASSETTE (ABC) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AF133903; AAF14372.1; -.
DR EMBL; AF186585; AAD56419.1; -.
DR MGD; MGI:1351619; Abcb11.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transmembrane; Transport.
DR D0WAIN 1 62
FT D0WAIN 63 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 84 147 EXTRACELLULAR (POTENTIAL).
FT D0WAIN 148 168
FT TRANSMEM 148 168
FT D0WAIN 169 215 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 216 236 POTENTIAL.
FT D0WAIN 237 240 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 241 261
FT D0WAIN 262 319 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 320 340
FT D0WAIN 341 353 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 354 374 POTENTIAL.
FT D0WAIN 375 755 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 756 776
```



```

FT DOMAIN 777 794 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 795 815 POTENTIAL.
FT DOMAIN 816 890 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 870 890 POTENTIAL.
FT TRANSSEM 891 911 POTENTIAL.
FT DOMAIN 912 979 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 980 1000 POTENTIAL.
FT DOMAIN 1001 1011 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 1012 1032 POTENTIAL.
FT DOMAIN 1033 1321 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 455 462 ATP (POTENTIAL).
FT NP_BIND 1113 1120 ATP (POTENTIAL).
FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 481 481 L -> P (IN REF. 2).
FT CONFLICT 633 633 T -> V (IN REF. 2).
SQ SEQUENCE 1321 AA; 146675 MW; 15B5EBF175D32967 CRC64;

Query Match 49.6%; Score 3213.5; DB 1; Length 1321;
Best Local Similarity 49.7%; Pred. No. 64e-176;
Matches 643; Conservative 237; Mismatches 376; Indels 37; Gaps 9;

QY 15 FWMGKSKKKEKE-KKPTVSTFAMPRYNWLDRLMYLGTMAAIIHGAAFLMMLVFG 73
DB 25 FHNDDKSKRLQDKKKGARVGFELFRFSSKDNMLWMSGVYCALLHGNMAQPGMIIVFG 84

QY 74 NMTDSFANAGISR-----NKTFFVI---INESTNNTOHQFINHLEEMTYAYY 119
DB 85 ILTDFIYEDIEQELISPGVKVMNNTIWMNSFNQNTNGTSCGLVDINSEVIFSGI 144

QY 120 YSGIGAGVLVAAYIQVSWFCLAAGROILKIRKOFHAIIMROEIGWEDVDHVGELNRLTD 179
DB 145 YAGVAVALLIGVQIRLWITGARQIRKMRKFFRIMRKEIGWEDCTSVSGELNRSFD 204

QY 180 DVSKINGIGDKIMGFQSFQSTATFTFTGIVGTRGWKLTILVLAISVPLGLSAAIAWAKILS 239
DB 205 DINKIDEAIDOMALFLORLSTALSGLLGLGYRGWKLTVILAVSPGIGAAVIGLSVA 264

QY 240 SFTDKELLAYAKAGAAVEELAIRIVIAFGQKKKLELRYKNLEAKGIGIKKAITANI 299
DB 265 KFTLELKAAGAGSIADEVLSSIRTVAAFGGENKEVEREYKMLFAQRMGIWKGVMGE 324

QY 300 SIGAAFLIYASALAFWYGTSLVL-SSEYSIGVLTFTVFSVLIGAFSGIQAQSPSTEAPA 358
DB 325 FTGYMNCILFFCALAFWYGSRLVDEGETPTGLIQIFCLVIAAMNIGNASSCLEIFS 384

QY 359 NARGAAVEPKIIDNKPSIDSYSGSKGHPDNTKGNLEFKNVHPSYSPRKEVKILGLNLK 418
DB 385 TGSAASSIFQITDROPMVDMCDMSGDGYKLDRIKGETEFHNVTFFHPSRPEVKILNLSMV 444

QY 419 VQSGQTVALVNGSGCKSTTVOLMORLYDPTDGMVCDIGQDITINVHLREITGVVSQ 478
DB 445 IKGETTAFVSSGAGKSTALQITQIFDPCGWNVLGDHDKSLRNLWRDQIGIVEOE 504

QY 479 PVLFTATINIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDPLVGERGAOLSGGQK 538
DB 505 PVLFTTIAENIRLGREETMEDIVQAADANAYNEIMALPOQFDPLVGEQSGGQSGQK 564

QY 539 ORTAIRALVRNPKILLDEATSALDTESEAVVQVQALDKARKGRTTIVIAHRLSTVRNAD 598
DB 565 QRYAIARALKPKILLDMATSDALDNESEAKVQOALNKIQHGTITISVAHRLSTVRSAD 624

QY 599 VIAFGDGVIVEGNDELKMGKGYFLKVTWQTRGNETELENA-TGESKSSSDALEM-- 655
DB 625 VIIGFEGHTAVERGTHEELERKGVYFVMTLQSQEDNTHKTGTGKDKTDEGTPERTF 684

QY 656 ---SPKDSGSLIKRRSTR-----SIHAPQODGRKLGTKEDLNENVPVPSW 700
DB 685 SRGSYQDSLRASTRQSKQLSHLSHEPPLAIGDHSSEYEDRK--DNDVLVEVEFAPVR 742
QY 701 RILKLNSTWPFVVGIFCAIINGGLQPAFSIFRSKIIGIFTRDEDPETKQNSNMFSYL 760

```

```

DB 743 RILKYNISEWPYILVGCALCAANGAVTPIYSLLFSOILKTFSL-VDKEQORSEIYSMCLF 801
QY 761 FLVLGIISFITPFLQGTFCAGKEILTKRLRYMVFMSMLRQDVSFDDPKNTTGTALTTRL 820
DB 802 FVILGCVSLTQFLQGYNEAKSGELTKRLRKFGEKAMLRQDQIGWFDLKNNGPVLTTRL 861
QY 821 ANDAAQVKAIGSRLLAVITONTANLGTGIIISLIWGQTLTLLLAIVPIATIAVGVEMKM 880
DB 862 ATDASQVQATGSGVGMWNSFTNIFVAVLIAFLFNWKLISLVISVFFFLALSAGAVQTKM 921
QY 881 LSGQALKDKKELEGAGKIAEINENFRVTVSLTREQKFEYMAQSLQVYPRNSLRKAHIF 940
DB 922 LTGFASQDKLELKAQITNEALSNIPTVAGIGVEGRFKAPEVELEKSYKTAIRKANVY 981
QY 941 GVSFSTQAMMYFSYAGCFRFGAYLVANFEMFQDVLVLSAIVFGAMAVQVSSFADPY 1000
DB 982 GLCYAFSGISFLANSAAYRYGYLIVEDLNFYSVFRVSVSSIAMSATAVGRTFSYTPSY 1041
QY 1001 AKAKYSAAHVIMIEKSPILDSYSPHGLKPNTELEGNTFNEVNYPTRPDPVQLGSL 1060
DB 1042 AKAKISAAFFQLDKPKPIDVSGAGEKWNDFQKIDFIDCKFTYPSRPDIQVLNGLSV 1101
QY 1061 EYKKGQTLALVSSGCGKSTVYQLLERFYDPLAGSVLIDGKEIKHLNQLWLAHLGIVSQ 1120
DB 1102 SVDPGQTLAFVSSGCGKSTQLLERFYDPPQGVIMIDGHDSKKVNVQFLRSNIGIVSQ 1161
QY 1121 EPILEDCAENAIYAGDSRVVSHHEIMOAKENIHHFIETPLPEKYNTRVGDKGTQLSG 1180
DB 1162 EPIVLEDCSITMDNIKYGDNTEKISVERATAAAQALHDFVMSLPEKYEYTNVGIQGSLSR 1221
QY 1181 GOKORTAIRALVRPHILLDEATSEKTSALDTESEKVVQVQALDKAREGRICIVIAHRLSTIQ 1240
DB 1222 GEKORTAIRALVRPHILLDEATSEKTSALDTESEKVVQVQALDKAREGRICIVIAHRLSTIQ 1281
QY 1241 NADLVVWFONGVKGHGTQQLLAQKGIYFSMV 1273
DB 1282 NSDIIAVMSQGVVIEKGTGKMLMDQKGYKLV 1314

RESULT 14
AB11_RAT
ID AB11_RAT STANDARD; PROT; 1321 AA.
AC O70127;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11)
DE (Sister of P-glycoprotein).
GN ABCB11 OR BSEP OR SPGP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=98212048; PubMed=9545351;
RA Gerloff T., Stieger B., Hagenbuch B., Madon J., Landmann L., Roth J.,
RA Hofmann A.F., Meier P.J.;
RT "The sister of P-glycoprotein represents the canalicular bile salt
RL J. Biol. Chem. 273:10046-10050(1998).
CC -!- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
CC -!- INTO THE CANALICULUS OF HEPATOCYTES.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
CC MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
CC SITU.
CC -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.

```

1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U69487; AAC40084.1; -
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmemb.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
KW ATP-binding; Transmembrane; Transport.
FT DOMAIN 1 62
FT TRANSMEM 63 83
FT DOMAIN 84 147
FT TRANSMEM 148 168
FT DOMAIN 169 215
FT TRANSMEM 216 236
FT DOMAIN 237 240
FT TRANSMEM 241 261
FT DOMAIN 262 319
FT TRANSMEM 320 340
FT DOMAIN 341 353
FT TRANSMEM 354 374
FT DOMAIN 375 755
FT TRANSMEM 756 776
FT DOMAIN 777 794
FT TRANSMEM 795 815
FT DOMAIN 816 869
FT TRANSMEM 870 890
FT DOMAIN 891 911
FT TRANSMEM 912 979
FT DOMAIN 980 1000
FT TRANSMEM 1001 1011
FT TRANSMEM 1012 1032
FT DOMAIN 1033 1321
FT NP_BIND 455 462
FT ATP (POTENTIAL).
FT NP_BIND 1113 1120
FT CARBOHYD 109 109
FT CARBOHYD 116 116
FT CARBOHYD 122 122
FT CARBOHYD 125 125
FT SEQUENCE 1321 AA; 146257 MW; 5443F4EF7B9FB1F6 CRC64;
Query Match 49.6%; Score 3211.5; DB 1; Length 1321;
Best Local Similarity 50.3%; Pred. No. 8.3e-176;
Matches 648; Conservative 234; Mismatches 370; Indels 35; Gaps 11;
QY 20 KSKKKKEK-KERRPTVSTFAMFYSNWLDRMLVGTMAAIIHGAALPLMLVFGNMTDS 78
DB 30 KKSRLQKKKEGDIRVGFELFSSSKDIWLMMLGGVCALLHGMAGPQGLIIFGIMTDI 89
QY 79 FANAGISR-----NKTFPVINES-----TNNTQPHINLEEMTYAYYSGI 123
DB 90 FIKYIERQELEIPGRACVNNITI-VWINSFQHNNTGTVCGLVDISESEMIKFSIYAGV 148
QY 124 GAGVLVAAYIOVSFWCLAGROLIKTKQFEHAIHQEIGFVDHVDGELNRLTDDYSK 183
DB 149 GMTVLIIIGYFQIRLWITGARGQIRRMKRIYFRIMRKEIGWEDCTSVGELNSRFADDIEK 208
QY 184 INEGIGDKMGFFQSIATFTTGFIVGFTRGWKLTVILAIISVPLGSLAAIWAAILSSFTD 243
DB 209 INDIAIQLAHLFLQRMSTAMCGLLGLGYRGWKLTVILAVSPILIGIAVGLSIKATE 268
QY 244 KELLAYAKAGAAVEVLAIRTVIAPGGQKKELERYKNLEAKGIGIKKAITANISICA 303

DB 269 LELKAYAKAGASIADEVLSSTRTVAAFGGENKEVEREKNLVFAQRWGIWKGMVMEFTFY 328
QY 304 AFLIIIVASYALAEWYGTSLV--SSEYSIGOVLVFFSVLIGAFSIGQASPSIEAFANARG 362
DB 329 MWCLIFCYALAEWYGTSLVDEEYTPGTLVQIFLCVLAAMNIGHASSCLEIFSTGCS 388
QY 363 AAYEIFKIIDNKPSIDSYSKSGHKPNIKGNLEFKNVHFSYPSRKEVKILKGLNKLKQSG 422
DB 389 AATNIFQIDRQPVIDCMGSGDYKLDRIKGEIEFHNVTTHYPSRPDVKILDNLNLSWKFG 448
QY 423 QTVALVGNSCGKSTTVQLMORLYDPTDGMVCIDGODIRTNVHRLREITGVVVSQBPVL 482
DB 449 ETTALVSGSGAGKSTALQIQRFDYDCEGMVTLDDGHDIRSLNRLWLDQIGIVEQBPVL 508
QY 483 ATTIAENIRYGRNVMTDETEKAVKEANAYDFIMKLPNKFDTLIVGERGAQLSGQKORIA 542
DB 509 STTIAENIRGREDATMEDIVQAKADANAYFMALPQPDFTLVGGGQMSGQKORVA 568
QY 543 IARALVRNPKILLDEATSALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAG 602
DB 569 IARALIRNPKILLDMATSDALDNESEARVOEALNKIQHGTIIISVAHRLSTVRAADVI 628
QY 603 FDDGVIVKGNHDELMEKEGIYKLYVTMOTRGNIELENA-TGESKESDALEM-----S 656
DB 629 FEHGVAVERTHEELLERKGVYFMLVTLQSGDNAKETSIMGKDATEGGTGLTERTSRGS 688
QY 657 PKDGSGLIKRRSTR---SIHAP--QGDRKLGKTED-----LNENVPVPSVWRILKLN 706
DB 689 YRDSLRASIQRSKSQLSLTHDPPPLAVADHKSYKDSKNDVLVEVEPAPVRRILKYN 748
QY 707 STEMPYVWVGIFCAIINGLOPAPFISRIIGIFTRDEPETKRONSMFSLVFLVGI 766
DB 749 IPEWHYILVGLSAAIINGAVTPIYSLFSLQLGTFSI-LDKEQQRSEIHSKMLFFVLGC 807
QY 767 ISFTTFLOGFTCKAGEILTKRLYVVFMSMLRQDVSWDFDPKNTGALTTLANDAAQ 826
DB 808 VSIPTFQLOGYTFKSGELLTKRLKFKGFRAMLCQDIGNFDDLRNPNPGLVTLRLATDASQ 867
QY 827 VKGAIGSLAVITONIANLGTIIISLYGWLTLALLAIVPIIAIAGVVMKMLSQAL 886
DB 868 VQAGATSGQVMVNSFTNIIAALLIAFFSWKLSLIITIFFFLALSGAVQTKMLTGFS 927
QY 887 KDKLEGAQKATEATENFTVVSLETRQKFFVYQAQSLQVYRNSLRKAHIFGVYSFI 946
DB 928 QDKQALEKAGQITSEALSNITVAGIGVEGRFKAFVELELQTSYKTAVRKANIIJLCF 987
QY 947 TQAMMYFSYAGCPFRGAYLVANEFMNFQDVLVFSALVFGAMAVGVSSFAPDYAKAVS 1006
DB 988 SQGIAPLANSAYRYGGLIAYEGLGFHSHVFRVVSVALSATAVGRTFTSPYAKAKIS 1047
QY 1007 AAHVIMIIEKSPILDSYSPHGLKPTNLTGKNTFNEVFNFTPTDIPVLOGLSLEVKKQ 1066
DB 1048 AARFQLDLDRKPPINVYSEAGEKWDNFQKIDFIDCKFTYPSRPIQVNLNGLSVSNPQG 1107
QY 1067 TLALVSGSGCKSTVOLLERFYDPLAGSLVDGKELKHLNVQWLAHLGIVSOEPLTFD 1126
DB 1108 TLAFFVSGGCKSTSIQLLERFYDQDGTMDHGDHKKVNIQFLRNSNIGIVSOEPLTFD 1167
QY 1127 CSAENIAYGNSRWVSHHEEIMQAANEKANIHFETIPEKYNTRVGDKGTQLSGQKQRI 1186
DB 1168 CSIMDNKYDNTKEISVERAIAAKQAQLHDFVMSLPEKYETNVGIGQSLSRGEKQRI 1227
QY 1187 AIARALVRQPHILLDEATSALDTESEKVVQVQALDKAREGRTCTIVIAHRLSTONADLI 1246
DB 1228 AIARAIVRDPKILLDEATSALDTESEKVTQALDKAREGRTCTIVIAHRLSTONSDIIA 1287
QY 1247 VFQNGKVEKTHQOALLAQKGIYFSMV 1273
DB 1288 VVSQGVIEKGTHEKLMQAQKGIYKLV 1314
RESULT 15


```
QY 61 HGAALPLMLLVFGNMTDSFANAGISRNKTFPVILINESITNNTOHFINHLEEBMTYYAYY 120
Db 60 HGAALPLMLLVFGNMTDSFANAGISRNKTFPVILINESITNNTOHFINHLEEBMTYYAYY 119
QY 121 SGIGAGVLVAAYIQVSWFCLAAAGRQILKIRKOFFHAIMRQEIIGWDFDVGELNTRLTDD 180
Db 120 SGIGAGVLVAAYIQVSWFCLAAAGRQILKIRKOFFHAIMRQEIIGWDFDVGELNTRLTDD 179
QY 181 VSKINGIGDKIGMFTQSTATFTTGIVGFTGRGWKLTILVILAIISPVLGLSAAIWAAILSS 240
Db 180 VSKINEGIGDKVGMFTQSTATFTTGIVGFTPCWKLTILVILAIISPVLGLSAAIWAAILSS 239
QY 241 FTDKELLAYAKAGAAVEEVLAAIRTVIAFGGKKELERYKNLEAKGIGIKKAITANIS 300
Db 240 FTDKELLAYAKAGAAVEEVLAAIRTVIAFGGKKELERYKNLEAKGIGIKKAITANIS 299
QY 301 IGAFAFLLIYASALAFWYGTSLVLSSEYSIGQVLTIVFFSVLIGAFSIGQASPSIEAFANA 360
Db 300 IGAFAFLLIYASALAFWYGTSLVLSSEYITIGVLTIVFFSVLIGAFSIGQASPSIEAFANA 359
QY 361 RGAAYEFKLIIDNKPSTDSYKSGHKPDNIKGNLEPKNVHFPSPKVKILKGLNKKVQ 420
Db 360 RGAAYEFKLIIDNKPSTDSYKSGHKPDNIKGNLEPKNVHFPSPKVKILKGLNKKVQ 419
QY 421 SGOTVALVNSGCGKSTTVQLMQRLYDPTDGMVCIQDQDIRTINVRLREITGVYSQEPV 480
Db 420 SGOTVALVNSGCGKSTTVQLMQRLYDPTDGMVCIQDQDIRTINVRLREITGVYSQEPV 479
QY 481 LFATTIAENTRYGRENVTWDEIEKAVKEANAYDFIMKLPNKEDTTLVGERRAQLSGGQOKR 540
Db 480 LFATTIAENTRYGRENVTWDEIEKAVKEANAYDFIMKLPNKEDTTLVGERRAQLSGGQOKR 539
QY 541 IATARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
Db 540 IATARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 599
QY 601 AGFDDGVIVEKGNHDELMKEKGYFKLVMTMOTRNGEIELENATGESKESDALEMSPKDS 660
Db 600 AGFDDGVIVEKGNHDELMKEKGYFKLVMTMOTRNGEIELENATGESKESDALEMSPKDS 659
QY 661 GSSLIKRRSTRRSIHAPQGGDRKLGTCKEDLNENVPVSWRILKLNSTWEPYFVVGIFCA 720
Db 660 GSSLIKRRSTRRSIHAPQGGDRKLGTCKEDLNENVPVSWRILKLNSTWEPYFVVGIFCA 719
QY 721 IINGGLQPAFSIFSRIGITFTRDEDPETKRONSNFMSVFLVLGIIISFTTFPLOGFTFG 780
Db 720 IINGGLQPAFSIFSRIGITFTRDEDPETKRONSNFMSVFLVLGIIISFTTFPLOGFTFG 779
QY 781 KAGEILTCLRMYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSRLAVITQ 840
Db 780 KAGEILTCLRMYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSRLAVITQ 839
QY 841 NIANLGTGIIISLYGHWLTLLALLAIVPIITAIAGVEMKMLSGQALKDKKELEGACKIAT 900
Db 840 NIANLGTGIIISLYGHWLTLLALLAIVPIITAIAGVEMKMLSGQALKDKKELEGACKIAT 899
QY 901 EAIENFRTVSLTREQFEYMYAQSLOQVPYRNSLRKAHIFGVSFISITQAMMYESYAGCFR 960
Db 900 EAIENFRTVSLTREQFEYMYAQSLOQVPYRNSLRKAHIFGVSFISITQAMMYESYAGCFR 959
QY 961 FGAYLVANEFMNFODVLLVFSATVFGAMAVQVSSFPADYAKAKVSAAHVIMIIEKSPLI 1020
Db 960 FGAYLVANEFMNFODVLLVFSATVFGAMAVQVSSFPADYAKAKVSAAHVIMIIEKSPLI 1019
QY 1021 DSYSPHGLKNTLBNVTFFNEVFNYPTRPDIPVLOGLSLEVKKGTALVGSNGCGKST 1080
Db 1020 DSYSPHGLKNTLBNVTFFNEVFNYPTRPDIPVLOGLSLEVKKGTALVGSNGCGKST 1079
QY 1081 VVQLLERYFDPLAGSVLIDGKEIKHLNVQWMLRAHLGTVSQEPILFDCSIAENTAYGDNSR 1140
Db 1080 VVQLLERYFDPLAGSVLIDGKEIKHLNVQWMLRAHLGTVSQEPILFDCSIAENTAYGDNSR 1139
QY 1141 VVSHEEIMQAANEANIHHFIETLPEKYNTRVGDGKGTOLSGGQKORIAIARALVRQPHILL 1200
```

```
Db 1140 VVSHEEIMQAANEANIHHFIETLPEKYNTRVGDGKGTOLSGGQKORIAIARALVRQPHILL 1199
QY 1201 LDEATSALDTESEKVOEALDKAREGRTCVIAHRLSTIONADLIVVFQNGKVKHEGTHQ 1260
Db 1200 LDEATSALDTESEKVOEALDKAREGRTCVIAHRLSTIONADLIVVFQNGKVKHEGTHQ 1259
QY 1261 QLLAQKGIYFSMVSVOAGAKR 1281
Db 1260 QLLAQKGIYFSMVSVOAGAKR 1280
RESULT 2
O02793
ID O02793 PRELIMINARY; PRT: 1285 AA.
AC O02793;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN-1.
GN MDR1.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Longley M., Crawford A.M.;
RT "Ovine mdr1 gene.";
RC Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RL -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; U78609; AAB58489.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP-A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ
SEQUENCE 1285 AA; 142020 MW; 90153B617C44856F CRC64;
Query Match 87.8%; Score 5685.5; DB 6; Length 1285;
Best Local Similarity 87.4%; Pred. No. 0;
Matches 1123; Conservative 79; Mismatches 78; Indels 5; Gaps 3;
QY 1 MDPEGRKG-SAEKNFWKMKK---SKKKEKKKPTVSTFAMFYSNWLDRLYMLVGTM 56
Db 1 MDLEGRNGRAAGGNFLKRDKKRFFSKKDEKKRPTVSTFMTFYSNWLDRLCMLVGTL 60
QY 57 AAIHGAALPLMLLVFGNMTDSFANAGISRNKTFPVILINESITNNTOHFINHLEEBMTTY 116
Db 61 AAIHGAALPLMLLVFGDWTDSFAGAGNLGNTILSNISNTSTIDRTE-YGKKEKEMTY 119
QY 117 AYYSGIGAGVLVAAYIQVSWFCLAAAGRQILKIRKOFFHAIMRQEIIGWDFDVGELNTR 176
Db 120 AYYSGIGAGVLVAAYIQVSWFCLAAAGRQVHRIRKOFFHAIMRQEIIGWDFDVGELNTR 179
QY 177 LTDDYSKINEGIDKIGMFFQSIATFTTGIVGFTGRGWKLTILVILAIISPVLGLSAAIWA 236
Db 180 LTFNDSKINEGIDKIGMFFQSIATFTTGIVGFTGRGWKLTILVILAIISPVLGLSAAIWA 239
QY 237 ILSFTDKELLAYAKAGAAVEEVLAAIRTVIAFGGKKELERYKNLEAKGIGIKKAIT 296
Db 240 ILSFTDKELLAYAKAGAAVEEVLAAIRTVIAFGGKKELERYKNLEAKGIGIKKAIT 299
QY 297 ANISITGAAPLLIYASALAFWYGTSLVLSSEYSIGQVLTIVFFSVLIGAFSIGQASPSIEA 356
Db 300 ANISITGAAPLLIYASALAFWYGTSLVLSREYSIGQVLTIVFFSVLIGAFSIGQASPSIEA 359
QY 357 FANARCAAYEIFKIIDNKPSTDSYKSGHKPDNIKGNLEPKNVHFPSPKVKILKGLN 416
```

```
Db 360 FANARGAAYEFKIIDNKPDSISYNTGKHPDNKGNLEFRNVFHYPSRNEVKILGSLN 419
Qy 417 LKVGSGQVALVGNCGCKSTTVOLMORLYDPTDGMWCIDGQDITINVRHLREITGVVS 476
Db 420 LKVGSGQVALVGNCGCKSTTVOLMORLYDPTDGMWCIDGQDITINVRHLREITGVVS 479
Qy 477 QEPVLFAATIAENIRYGRNVTMDIEIKAVKEANAYDFIMKLPNKFDVLGGERGAQLSGG 536
Db 480 QEPVLFAATIAENIRYGRNVTMDIEIKAVKEANAYDFIMKLPNKFDVLGGERGAQLSGG 539
Qy 537 QKQRTAARALVRNPKILLDEATSEALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRN 596
Db 540 QKQRTAARALVRNPKILLDEATSEALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRN 599
Qy 597 ADVTAGFDGVTVEKGNHDELMKEGIYFKLVMTQTRGNELEENAYGKESSEDALEMS 656
Db 600 ADIAGLDGVTVEGSHDELMKGRGIYFKLVMTQTRGNELEENTGCELSKTYDDLYMS 659
Qy 657 PKDSSSLIKRRSTRSIHAPQGDQDKLGTREDLNENVPVPSFWRILKLNSTWPFYVVG 716
Db 660 SODSRSLIRKSTRSIRGSDQDKLSTETLDESVPVPSFWRILKLNITWPFYVVG 719
Qy 717 IFCAIINGLQAPSIPTSRIGITFTRDDETPETKRONSMFVLFLVGLIISFTTFELQG 776
Db 720 VFCAIINGLQAPSIPTSRIGITFTRDDETPETKRONSMFVLFLVGLIISFTTFELQG 779
Qy 777 FTEGKAGILTKRLRYMYFRSMLQDYSWFDPPKNTTCALTTRLANDAAQVKGAGISRLA 836
Db 780 FTEGKAGILTKRLRYMYFRSMLQDYSWFDPPKNTTCALTTRLANDAAQVKGAGISRLA 839
Qy 837 VITQNIANLGTGIIISLYGQWLILLAIIVPIIAIAGVVMKMLSGOALKDKKLELEGAG 896
Db 840 VITQNIANLGTGIIISLYGQWLILLAIIVPIIAIAGVVMKMLSGOALKDKKLELEGAG 899
Qy 897 KIATEAIENTFTVSLTREERFEYMAQSLQVPPYRNSLRKAHIFGVFSITQAMMYPSYA 956
Db 900 KIATEAIENTFTVSLTREERFEYMAQSLQVPPYRNSLRKAHIFGVFSITQAMMYPSYA 959
Qy 957 GCFRFGALVANEPMFQDVLVLSAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMLIEK 1016
Db 960 GCFRFGALVANEPMFQDVLVLSAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMLIEK 1019
Qy 1017 SPLIDSYSPHGLKPNLTLEGNTFNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSGGC 1076
Db 1020 IPLIDSYSTEGLKSTVEGSAFNDVFNYPTRPDIPVLQGLSLEVKKGOTLALVSGGC 1079
Qy 1077 GKSTVQVLLERFYDPLAGSLVDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSIAENTAYG 1136
Db 1080 GKSTVQVLLERFYDPLAGSLVDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSIAENTAYG 1139
Qy 1137 DNSRVVSHETMOAKKANIHHFTETLPEKYNTRVGDGKQTLGSGQKORTAIARALVRQP 1196
Db 1140 DNSRVVSHETMOAKKANIHHFTETLPEKYNTRVGDGKQTLGSGQKORTAIARALVRQP 1199
Qy 1197 HILLDEATSEALDTESEKVVQEALDKAREGRTICIVIAHRLSTIQONADLIIVVFQNGKVEH 1256
Db 1200 HILLDEATSEALDTESEKVVQEALDKAREGRTICIVIAHRLSTIQONADLIIVVFQNGKVEH 1259
Qy 1257 GTHQOLLAQKGIYFTMSVSVQAGAKR 1281
Db 1260 GTHQOLLAQKGIYFTMSVSVQAGAKR 1284
```

RESULT 3

```
Q9JK64 ID Q9JK64 PRELIMINARY; PRT: 1272 AA.
AC Q9JK64;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DR 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN 1A.
GN PGY1.
```

```
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=LIVER;
RA Hooiveld G.J.E.J., Wilms J.W.J., Hagenbuch B., Jansen P.L.M.,
RA Meijer D.K.F., Muller M.;
RT "Cloning and functional characterization of the rat multidrug
RT resistance protein Mr1a.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF257746; AAF69007.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SSQUENCE 1272 AA; 140326 MW; B5C1DF89DFC986C5 CRC64;

Query Match 87.0%; Score 5634; DB 11; Length 1272;
Best Local Similarity 86.5%; Pred. No. 1.3e-316;
Matches 1108; Conservative 80; Mismatches 83; Indels 10; Gaps 2;

Qy 1 MDPEGKGSAAENFNMKGKSKKKEKPTVSTFAMFRYSNWLDRMLVGTMAAII 60
Db 1 MELEEDLNGRADNFKSMGKKS-KKEKKEKPAVSULTMFRYAGWLDREYMLLGLTAAII 59
Qy 61 HGAALPLMLVFGNMTDSPANAGISRNKTFPVIINESIINTQHFNIHLEEMTYAYY 120
Db 60 HGIALPLMLVFGDMTDSFANVGNRSMF-----YNATDIYAKLEDEMTYAYY 110
Qy 121 SGIGAGVLVAAYIQVSWFCLAAGRQILTKRQFFHAIMRQEGWFDVHDVGEINLRLTD 180
Db 111 TGIGAGVLVIAYIQVSWLWCLAAGRQILTKRQFFHAIMRQEGWFDVHDVGEINLRLTD 170
Qy 181 VSKINIGDKIGMFTQSIATFTFTGWTGLTILAIISPVLGSAATWAKILSS 240
Db 171 VSKINIGDKIGMFTQAMATFEGGIIGTRGKTLTILAIISPVLGSAWAKILSS 230
Qy 241 FTDKELLAYAKAGAAVEVLAAITRTVIAFGQKKELRYKNLEAKGIGIKAITANIS 300
Db 231 FTDKELOAYAKAGAAVEVLAAITRTVIAFGQKKELRYNNLEAKRIGIKAITANIS 290
Qy 301 IGAALFLIYASALAFWYGTSLVLSSEYSTIGVLTFFSVLIGAFSIGQASPSIEAFANA 360
Db 291 MGAALFLIYASALAFWYGTSLVISKEYTIGVLTFFSVLIGAFSVGQASPNIEAFANA 350
Qy 361 RGAAYEIFKIIDNKPDSISYSGHGXPDNKNLEFNKYNHVSYPYSKVEYKILGKLNKVK 420
Db 351 RGAAYEVFSTIDNKPDSISYSGHGXPDNKNLEFNKYNHVSYPYSKVDQVILGKLNKVK 410
Qy 421 SGQTVLVGNSGGCKSTTVQLMORLYDPTDGMWCIDGQDITINVRHLREITGVVSQEPV 480
Db 411 SGQTVLVGNSGGCKSTTVQLMORLYDPTDGMWCIDGQDITINVRHLREITGVVSQEPV 470
Qy 481 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDVLGGERGAQLSGGQKOR 540
Db 471 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDVLGGERGAQLSGGQKOR 530
Qy 541 IATARALVRNPKILLDEATSEALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
Db 531 IATARALVRNPKILLDEATSEALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 590
Qy 601 AGFDDGVIVKGNHDELMKEGIYFKLVMTQTRGNELEENAYGKESSEDALEMSPKDS 660
Db 591 AGFDDGVIVKGNHDELMKEGIYFKLVMTQTRGNELEENAYGKESSEDALEMSPKDS 650
```



```
QY 1141 VVSHEEIMAAKEANIIHFIETLPEKYNTRYGDKGTQSGGOKORIAIARALVROPHILL 1200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1131 VVSHEEIVKAAKEANIHOFIDSLPEKYNTRYDRTGTQSGGOKORIAIARALVROPHILL 1190
QY 1201 LDEATSDLTSEKVVQVQALDKAREGRTICIVIAHRLSTIONADLIVVFQNGKVEHGTQ 1260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1191 LDEATSDLTSEKVVQVQALDKAREGRTICIVIAHRLSTIONADLIVVFQNGKVEHGTQ 1250
QY 1261 QLLAQKGIYFSMVSVQAGAKR 1281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1251 PLLAQKGIYFSMVSVQAGAKR 1271

RESULT 5
Q9TSU2
ID Q9TSU2 PRELIMINARY; PRT; 1163 AA.
AC Q9TSU2;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DE 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE MULTI-DRUG RESISTANCE RELATED (FRAGMENT).
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
SEQUENCE FROM N.A.
RA Okai Y., Nakamura N., Matsusiro H., Kato H., Setoguchi A., Yazawa M.,
RA Okuda M., Watari T., Hasegawa A., Tsujimoto H.;
RT "Molecular analysis of multidrug resistance in lymphoma cells in the
RT cat.";
RL Am. J. Vet. Res. 0:0-0(1999).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AB029153; BA87071.1;
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
FT NON_TER 1
FT TER 1163
SQ SEQUENCE 1163 AA; 128511 MW; 1B5B413776A93A26 CRC64;

Query Match 85.3%; Score 5521.5; DB 6; Length 1163;
Best Local Similarity 94.2%; Pred. No. 3.6e-310;
Matches 1097; Conservative 34; Mismatches 32; Indels 1; Gaps 1;

QY 61 HGAALPLMLVFGNMTDSFANAGISRNKTPPVLIINESITNNTTOHFFINHLEEMTTAYYY 120
Db 1 HGAALPLMLVFGNMTDSFANAGISRNLTTLNITGESIVNDS-YFINRLEEMTTAYYY 59
QY 121 SGIGAGVLVAAYIQVSWFLAAGRIKIRKQFFHAIMRQIEGWDFVDVHVGELNTRLTDD 180
Db 60 SGIGAGVLVAAYIQVSWFLAAGRIKIRKQFFHAIMRQIEGWDFVDVHVGELNTRLTDD 119
QY 181 VSKINEGIGKIGMFQSIATFTFTGTVGTRGWKTLTVILATSPVLGSAATWAKILSS 240
Db 120 VSKINEGIGKIGMFQSMATFTFTGTVGTRGWKTLTVILATSPVLGSAATWAKILSS 179
QY 241 FTKELLAYAKAGAAVEVLAARTVIAFGQKKELERYKNLEAKGIGKAITANIS 300
Db 180 FTKELLAYAKAGAAVEVLAARTVIAFGQKKELERYKNLEAKGIGKAITANIS 239
QY 301 IGAFLLIYASALAFWYGTSLVLSSEYSIGQVLTTFVFSVLICAFSPISGASIEAFANA 360
Db 240 IGAFLLIYASALAFWYGTSLVLSSEYSIGQVLTTFVFSVLICAFSPISGASIEAFANA 299
QY 361 RGAAYEIFKIIDNKPSIDTSYKNGHPRDNKIGNLEFKNHFVSPSRKEVKILKGLNLKQV 420
```

```
Db 300 RGAAYEIFKIIDNKPSIDTSYKNGHPRDNKIGNLEFKNHFVSPSRKEVKILKGLNLKQV 359
QY 421 SGQTVALVNGSGCGKSTTVQLMORLYDPTDGMVCIDQDQIRTNVRLREITGVVSQEPV 480
Db 360 SGQTVALVNGSGCGKSTTVQLMORLYDPTDGMVSDQDQIRTNVRLREITGVVSQEPV 419
QY 481 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGEKAQLSGGOKOR 540
Db 420 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGEKAQLSGGOKOR 479
QY 541 IAIARALVRNPKILLDEATSDLTSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADVI 600
Db 480 IAIARALVRNPKILLDEATSDLTSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADVI 539
QY 601 AGFDDGVIVKEGNHDELMKEGIYFKLVMTQTRNGNIELENATGESKSSDALEMSPKDS 660
Db 540 AGFDDGVIVKEGNHDELMKEGIYFKLVMTQTRNGNIELENATGESKSSDALEMSPKDS 599
QY 661 GSSLIKRRSTRRSIHAPQODRKLTGKEDLNENVPVSWRILKLNSTWMPVYVVGIFCA 720
Db 600 GSSLIKRRSTRKSIHASQODRKLTGKEDLNENVPVSWRILKLNSTWMPVYVVGIFCA 659
QY 721 IINGLQPAFSIIFRSRIIGITRDEDPETKRONSMFSLFLVLGIIISITFFLQGTFFG 780
Db 660 IINGLQPAFSVILSRIIGITRDEDPETKRONSMFSLFLVLGIIISITFFLQGTFFG 719
QY 781 KAGEILTKRLRYMFRSMLRQDVSNFDDPKNTGALTTRLANDAAQVKAIGSLRATVITQ 840
Db 720 KAGEILTKRLRYMFRSMLRQDVSNFDDPKNTGALTTRLANDAAQVKAIGSLRATVITQ 779
QY 841 NTANLGTGIIISLIYQWLTLLLAIVPIIAIAGVYVEMKMLSGQALKDKKELEGAGKIAT 900
Db 780 NTANLGTGIIISLIYQWLTLLLAIVPIIAIAGVYVEMKMLSGQALKDKKELEGAGKIAT 839
QY 901 EAIENFRTVVSILTREQKFEYMAQSLQVYPYRSLRKAHIFGVFSFISITQAMMYFSYAGCFR 960
Db 840 EAIENFRTVVSILTREQKFEYMAQSLQVYPYRSLRKAHIFGVFSFISITQAMMYFSYAGCFR 899
QY 961 FGAYLVANEFMFQDVLVLFSAIVFGAMAVGQVSSFADPYAKAKYSAAHVIMIEKSPLI 1020
Db 900 FGAYLVANEFMFQDVLVLFSAIVFGAMAVGQVSSFADPYAKAKYSAAHVIMIEKSPLI 959
QY 1021 DSVSPHGLKPNLTLEGNTVENEVFNYPTRDPVPLQGLSLEVKGTQALVSGSGCGKST 1080
Db 960 DSVSPHGLKPNLTLEGNTVENEVFNYPTRDPVPLQGLSLEVKGTQALVSGSGCGKST 1019
QY 1081 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHLGIYVQEPILFDCSIAENIAYGDSNR 1140
Db 1020 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHLGIYVQEPILFDCSIAENIAYGDSNR 1079
QY 1141 VVSHEEIMAAKEANIIHFIETLPEKYNTRYGDKGTQSGGOKORIAIARALVROPHILL 1200
Db 1080 VVSHEEIMAAKEANIIHFIETLPEKYNTRYGDKGTQSGGOKORIAIARALVROPHILL 1139
QY 1201 LDEATSDLTSEKVVQVQALDKAR 1224
Db 1140 LDEATSDLTSEKVVQVQALDKAR 1163

RESULT 6
Q60502
ID Q60502 PRELIMINARY; PRT; 1169 AA.
AC Q60502;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DE 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE P-GLYCOPROTEIN.
GN PGP-1.
OS Cricetulus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
```


Db 368 ARGAAVTIENIIDNQPKIDFSKEGLKDPDKIKGDIIEKNVIFTYPSRKDIQVLKGLNLNI 427
QY 420 QSGQTVLVNGSGCKSTTVQLMQRLYDPTDGMWIDGODITINVRHLREITGVVSGEP 479
Db 428 PSQKTVLVNGSGCKSTTVQLIQRFYDPDGVITLDGODIRSLNIRYLREIGVSGEP 487
QY 480 VLPFATTIAENIRYGRNVTMDIEKAEKANAYDFIMKLPNKFDTLVGERGAOLSGGQK 539
Db 488 ILFDFTDIADIRYGRDVTKEIERATKANAYDFIMKLPNKFDTLVGERGTQLSGGQK 547
QY 540 RTAIRALVRNPKILLDDEATSALDTESEAVVQVADLKARKGRTTIVIAHRLSTVRNADV 599
Db 548 RTAIRALVRNPKILLDDEATSALDTESEAVVQVADLKARKGRTTIVIAHRLSTVRNADV 607
QY 600 IAGFDGVTIVEKGNHDELMKEGIYFKLVYTMOT----RCNTELENAAGESSESDALEM 655
Db 608 IAGFDNGVIVEGSHKELMERGVFNVLVTQVETSKDTEEDLETHIYEKK-----I 660
QY 656 SPKDGSSLIKRRSTRRSIHA--POQODRKLTGKE--DLNENVPVPSFWRIKLKLNSTWEPY 712
Db 661 PVYTHSNLVRKSSRNTIKSKVPETEDKEVDEEKKKEEGPPVPSFFKMKLNKPWEPY 720
QY 713 FVVGIFCAIINGLOPAPSIISRIIGIFTRDEDPETK--QNSNMFSLVFLVGLIISFIT 771
Db 721 FVVGIVCAMILATOPAPAFIIFRSIIGVFA--GPVSQMRSESSMYSLFLALGVGSFIT 777
QY 772 FFLQGFTECKAGEILTFLRLRYVVFERSMLRQDYSWFDDPKNTTGALTTRLANDAAOVKGA 831
Db 778 FFLQGFTECKAGEILTFLRLYLSGFSKMLRQEGWFDSDKNSGALTTRLATDASOVQAT 837
QY 832 GSRVAITONIANLGTGIIISLYIQWLTLLLAIVPIITAIAGVEMKMLSQALKDKKE 891
Db 838 GTRALLAQNANLGTGIIISLYIQWLTLLLAIVPPIITAIAGVEMKMLSQALKDKKE 897
QY 892 LEGAGKATEAINTFTVSVLSTREQFEYMYAQSLQVYPRNSLRKAHIFGVSFSTQAMM 951
Db 898 LERAGKISTDAVINRTVSVLSTRERFEAMYKSLSEGPYRNSIKKAHLHGLTYGLSQAH 957
QY 952 YFSYAGCFR--FGAYLVANEFMNFQDVLVFSALVFAMAVGVQVSPFADYAKAKVSAAHV 1010
Db 958 VLCLCWVFSVLGAYLVVEGLMKLDEVFLVSSAIVLGAMALGOTSFPADYTKAMISAHH 1017
QY 1011 IMIIEKSPLIDSVSGHGLKPNTELEGNVTNEVFNTPRPDIPVLOGLSLEVKQGTAL 1070
Db 1018 FSLERVPQIDSYSDQGEKPKNSGNVPFKGVNFNTPRPDITVLQGLDISVKGQETLAL 1077
QY 1071 VGSNGCGKSTVOLLRRFYDPLAGSVLIDGKEIKHLNVQWLAHLGIYVSOEPIFDCSTA 1130
Db 1078 VGSNGCGKSTVOLLRRFYDPEGEVLVDGLSVRNLIQWVRAQMGIVSQEPILFDCSIG 1137
QY 1131 ENIAYGNSRVVSHHEIMOAKBANIHFETLPEKYNTRVGDGKGTQLSGGQKQRIATAR 1190
Db 1138 DNIAYGNNRKNVQTEBIEATAKEANIHSFTESLTDKYNTRVGDGKGTQLSGGQKQRIATAR 1197
QY 1191 ALVRQPHILLDDEATSALDTESEKVVQVADLKARKGRTTIVIAHRLSTTONADLIWVFON 1250
Db 1198 ALIRKPKILLDDEATSALDTESEKVVQVADLKARKGRTTIVIAHRLSTTONADLIWVFON 1257
QY 1251 GKVEKHTHOOLLAOKGIYFSMVSVQAG 1278
Db 1258 GKVEGHTHOOLLQLKGVFSVLTIQIG 1285

RESULT 9
Q90235
ID Q90235 PRELIMINARY; PRT; 1348 AA.
AC Q90235;
DT Q1-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE BLUE SALT EXPORT PUMP.
OS Raja erinacea (Little skate).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;

OC Elasmobranchii; Squala; Hypnosqualea; Pristioraja; Batoidea;
OX Rajiformes; Rajidae; Raja.
RN NCBI_TaxID=7782;
RP SEQUENCE FROM N.A.
RX MEDLINE=21340059; Pubmed=11447010;
RA Cai S.Y., Wang L., Ballatori N., Boyer J.L.;
RT "Blue salt export pump is highly conserved during vertebrate evolution
and its expression is inhibited by PFIC type II mutations.";
RL Am. J. Physiol. Gastrointest. Liver Physiol. 281:G316-G322(2001).
DR EMBL; AF367243; AAK52958.1; -.
SQ SEQUENCE 1348 AA; 148648 MW; 81906F641654CFDD CRC64;
Query Match 51.5%; Score 3331; DB 13; Length 1348;
Best Local Similarity 50.2%; Pred. No. 1.2e-183;
Matches 658; Conservative 250; Mismatches 340; Indels 62; Gaps 9;
QY 22 SKKKE-----KKEKKPTVSTFAMFRYSNLDRLYMLVGTMAAIIHGAALPLMKLVPGNM 75
Db 44 SKSDDNENLSKGDRIKIGFFQLFRFASCYEVLTMVFGVGCALLHGAQPAVLLVPGLL 103
QY 76 TDSFANAGISRNKTFPVIINESJTNNTOHFIN-----HLEEEETTYAY 118
Db 104 TDTFIADIELOELKDT--RKHCNNTIHWKNGSEYLNQHNETHSCGILLIENETMFAY 161
QY 119 YYSIGAGVLVAAYIQVSWCLAAAGROILKIRKOFFHAIMRQEIQWDFVHDVGEINLTRL 178
Db 162 YYVGIGCSVLLGYQFCFWTAAAROTQTKIRAYFQIMRMEMGWPCDNCVSGELATRMS 221
QY 179 DDVSKINEGIDKICMFPFQSTATFTFTGFTFGWKLTLVILAIISPLVGLSAIINAKIL 238
Db 222 DDINKINDAIDQVIGIFIQRFSTFVSGFLMGVNGWKLTLVIAVSPILGLRAALMALSV 281
QY 239 SSFTDKELLAYAKAGAAVEVLAIRTVIAPFGGOKKELERYKNLEAKGIGIKKAITAN 298
Db 282 ARLTGLKAYAKAGAAVEVLSIRIVAAFSGEKVEDRYDRLNLFVAQRKIRKIMG 341
QY 299 ISIGAFLIIYASVALAFWYGTSLV--SSEYSIGQVITVFSVLIGAFSGQASPSSTEF 357
Db 342 FFTGVYMMIIFCCVALAFWYGSKLVIQNEYTPGGLQVFLGVLVAAAMNLSQASPCLEAF 401
QY 358 ANARGAAVEIPIIDNPKSIDYSKSGHKPDNIKNGEKNVHFSYPSRKEVTKLGLNL 417
Db 402 ASGRGAVKIETIDREPIDCMSEGGYTLNKKVGDIEFHNVTNYPSPDVTDLRLSM 461
QY 418 KVQSGQTVLVNGSGCKSTTVQLMQRLYDPTDGMWIDGODITINVRHLREITGVVSG 477
Db 462 VIKAGETTAFVPGSGSGKSTAVELIQRFYDPKQGMVTLGDHDIRSLNIQWLSLIGIVEQ 521
QY 478 EPLVFATTIAENIRYGRNVTMDIEKAEKANAYDFIMKLPNKFDTLVGERGAOLSGGQ 537
Db 522 EPLVFSTTIEENIRYVREGLTKNDIVQAAKAEKANAYDFIMLPEKFNVLVGGGQMSGGQ 581
QY 538 KORTAIRALVRNPKILLDDEATSALDTESEAVVQVADLKARKGRTTIVIAHRLSTVRNA 597
Db 582 KORTAIRALVRNPKILLDDEATSALDNESEAIIVQGALDKVRFGRGRTTISIAHRLSTVRN 641
QY 598 DVIAGFDGVTIVEKGNHDELMKEGIYFKLVYTMOTRNEIELENAAGESSESDALEMSP 657
Db 642 DVIIGFEHGRAVERGRHAELELKERGIYFTLVLTQTEQALHEKA----RQVNGAIDEGA 697
QY 658 KDSGSSLIKRRSTRRSIHAPOGQDK-----LGTK----- 687
Db 698 SEK--QLLIRGSSRASVRSTRHQRSQVSEVLSDLSPGDVASAVRTPSISIGDEKDV 756
QY 688 EDLNNVPVPSFWRIKLKLNSTWEPYFVVGIFCAIINGLOPAPSIISRIIGIFTRDEDP 747
Db 757 EEEESIEAPVPSRIKLVNVSSEWPMFLGSLGAAVNGVNPYIALLFSQILGTESL-QNE 815
QY 748 ETKRONSNNFSLVFLVGLIISFITFFLQGTFFGKAGEILLTFLRYMYFRSMLRQDVSKED 807
Db 816 BEKINQINAICLFFVVGVLVSFLTQFLOSYPFAKSGELLTFLRLKRLGFOAMLRQEIWGED 875


```
QY 120 YSGIGAGVLVAAYIQVSWCLAAQRIKIRKQFFHAIHQEIGWFDVHDVGNLTRLTD 179
Db 118 YLGLGFAMFAYSIIQVCEWTFABRITHKLRKIYKAILRQIQSWFDIQQTGNLTARULT 177
QY 180 DYSKINEIGDKIGMFFOSIAFTFTGFTVTRGKWLFLVLAISPVVLGSLAAIWAIKULS 239
Db 178 DLERVREGLDKLSLFIQMVSFAFAGVCGFAYSWSMTLVMMVAPFIVISANWMSKIVA 237
QY 240 SFTDKELLAYAKAGAVEVLAARTVTIAPGOKKEKERYNKNLEEAAGIKKAIYANI 299
Db 238 TRTQVEQTYAVAGAIABETSSIRTVHSICGHRRETRFEALEKRGQTLGVLYFMVG 297
QY 300 SIGAFLLIYASALAFWYGSVLSS--EYSIGQVLVFFSVLIGAFSISQASPSIEAFA 358
Db 298 GVGFCQMCTVYSYALAFWGVLIINDPALDRGRIFTVFFAVMSGSAALGTCPLPHINTIS 357
QY 359 NARGAAVEIFKIIDNKPSIDSYSKSGHKPDNIKNLKFNHPSYPSRKEVKILGLNLK 418
Db 358 IARGAVRSVLGVINSRPRKIDPYSLDGIVLNNMGRSIRFNHFSYPSRRTLQILKGVSLQ 417
QY 419 VQSGQTVLGVNSCGKSTTVQLMORLYDPTDGMVCIDGODIRTINVRHLREITGVVSQE 478
Db 418 VSAGOKIALYVSSCGKSTNVNLLRFDYDTRGKVTIDIDVCDLNVQKLRQIGVVSQE 477
QY 479 PVLFTATIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFTDVLVBERGAQLSGQK 538
Db 478 PVLFDGTLFENIKMGYEQATMEEVQACRVANAADFTKRLPEGVGTTRVGERGVQLSGQK 537
QY 539 QRIAIARALNRPKILLDEATSDALDSEAVQVVALDKARKGRTTVIAHRLSTVRNAD 598
Db 538 QRIAIARAIKNPKILLDEATSDALDSEAVQVVALDKARKGRTTVIAHRLSTVRNAD 597
QY 599 VIAGFDGVIIVEKGNHDKERGIYFKLVMTQTRNGEIELENATGESKESDALEMSPK 658
Db 598 QIFVEKNGTIVEQTHAELANKRGVPEMTQAQVLQKEEVEVLDSDAESDVVSPDIALP 657
QY 659 DSGSLIKRSTRSIIHA-PQODRKLGTKEDLENVPPVSVFWRILKLNSTEWPYVVG 717
Db 658 HLSLSRKESKESISAVPSVRSMTQI-EMEDLRKPTMS--KIFYNRDKWGYFILGL 714
QY 718 FCAITINGLOPAFSIIFRSIRIIGFTIRDEDPETKRONSMF-SVLFLVLGLIISITFLOQ 776
Db 715 IACITGTVTPTAVLYAQIIQVIS---EPVQMKGHVLFWCAGFIVGLVHAFAPFFSA 771
QY 777 FTEGAGEILTKRLRYMVFNRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSRLA 836
Db 772 ICLGRGCEALTKLRFAPKLLRQNVGFYDDIRHGCTKLCRFAFDADPNVR-YVFTRLP 830
QY 837 VITQNIANLGTGIIISLIVGWLTLILLAIPIIATAGVVMKMLSGQALKDKKELEGAG 896
Db 831 GVLSSVVTIIGALVIGFIFGWQALILMVVPLIIGSGYFEMRQPFKKMRDTEILLEAG 890
QY 897 KIATEAENFRVTVSLTREKPFYMAQSLQVYPYRNSLRKAHFGVYSFISITQAMVFSYA 956
Db 891 KVASQAVENIRTVHALNRQEQHFMYCEYLKEYRENLCQAHYGVGFAPFSQSLFFMYA 950
QY 957 GCFRGAIVLANEFMNFQDVLVLSAIVFGAMAVGVSSFADYAKAKVSAAHVIMIEBK 1016
Db 951 VAFWIGAIFVDHNSMQPIDVYRFFAFMFCGMVGNISFIPDVVKARLAASILLFYLIH 1010
QY 1017 SPLIDSYPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKGQTTALVGSQC 1076
Db 1011 PSEIDNLSEGV-TTKISGHISFRNRYFNYPTRQIRVLRLGNLEINPGTITVALVGSQC 1069
QY 1077 GKSTVQVLLERFYDPLAGSVLIDGKEIKHLNVOWLRHAHGIYSQRPILFDCSIAENIAYG 1136
Db 1070 GKSTVALLERFYNQNKGVITVDGENIRNMNIRNLEQVCIVSQBPTFDCITMENICYG 1129
QY 1137 DNSRVVSHIEINQAAKEANHIFETLPEKYNTRYDGTQLSGGQKORIAIARALVRQP 1196
Db 1130 LDDPKPSYEQVVAARAKMANIHNVFLGLPEGYDTRVGEKGTQLSGGQKORIAIARALIRD 1189
QY 1197 HILLDEATSDALDSEKVVQVQALDKAREGRTCTIVIAHRLSTIQNADLIIVVFQNGKVKEH 1256
```

```
Db 1190 PILLDEATSDALDSEKIVODALEVARQGRCTCLVIAHRLSTIQSDVIMVQEGKATDR 1249
QY 1257 GTHQQLLAQKGIYFSMVSVQ 1276
Db 1250 GTHELLMKNDLYKRLCETQ 1269

RESULT 15
Q21349 ID Q21349 PRELIMINARY; PRT; 1321 AA.
AC Q21349;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE K08E7.9 PROTEIN.
GN K08E7.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
  investigating biology.";
RL Science 282:2012-2018(1998).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; 277666; CA801232.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1321 AA; 145062 MW; 68094121B30E8746 CRC64;

Query Match 42.4%; Score 2744; DB 5; Length 1321;
Best Local Similarity 45.1%; Pred. No. 9.3e-150;
Matches 585; Conservative 220; Mismatches 434; Indels 58; Gaps 12;

QY 21 KSKKKEKKPTVSTFAMERYSNWLDRLYLVTGTMALIHGAALPLMLVFNWTSFA 80
Db 47 KITRDAKEVNVKNSIPOLYRITTTLEKLLFGLVAVITGAGLPLMSILQGVSAFI 106
QY 81 NAGISRNKTFPIINETSNTNTHQFI-----NHLSEEMTTYYIYSGIGAGVLVAAY 132
Db 107 NE-----QIVN-----NNGSTFLPTGQNTKTFDEHDVNVVSVYAATVGMWAQ 154
QY 133 IQVSWCLAAQRIKIRKQFFHAIHQEIGWFDVHDVGNLTRLTDVSKINSGIDKI 192
Db 155 ITVTCYLVAVQOMNRLRREFVKISILRQEISWFDTNHSGTLATKLFNLERKVGEGDKI 214
QY 193 GMFFOSIAFTFTGFTVTRGKWLFLVLAISPVVLGSLAAIWAIKLSSFTDKELLAYAKA 252
Db 215 GMAFYLSQFTGFTVAFTHSWQLTVLMLAVTPIQALCGFAIKASMTFAIRFLRYAKA 274
QY 253 GAVAEVLAARTVTIAPGOKKEKERYNKNLEEAAGIKKAIYANISIGAAFLLIYASY 312
Db 275 GKVVEETISSIRTVVSLNGLRYELERYSTAVEEAKAGVLKGLFLGISFGAMQASNISF 334
QY 313 ALAFWYGTSLVLSSEYSIGQVLTVFFSVLIGAFSISQASPSIEAFANARGAAVEIFKIID 372
Db 335 ALAFYIGVGVHDSLNFQDMLTTFSSVMGSMALAGLPQALVLTQAQGAASGIYVELD 394
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 19:26:40 ; Search time 2230.96 Seconds
(without alignments)
7749.846 Million cell updates/sec

Title: US-09-672-725C-25

Perfect score: 6473

Sequence: 1 MDPEGGRKGSAAENFKWKGK.....LLAQKGIYFSWVSQVQAKR 1281

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09672725/runat_04112002_124342_18001/app_query.fasta_1.6435
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09672725.@CGL_1_1_7365.@runat_04112002_124342_18001 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NBG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vft:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1421	22.0	943	9	AL520322
2	1073.5	16.6	981	10	BF969667

3	1035	16.0	894	10	BF584668
4	1034	16.0	803	10	BG298756
5	1024	15.8	1019	10	BG248052
6	968.5	15.0	944	10	BF796582
7	958	14.8	851	10	BF969062
8	950	14.7	785	9	AV709991
9	916.5	14.2	795	10	BF313560
10	916	14.2	726	10	BG293345
11	910	14.1	587	10	BM053000
12	908.5	14.0	628	10	BG080311
13	889	13.7	563	10	BF964198
14	889	13.7	563	10	BF964218
15	889	13.7	563	10	BM052878
16	880	13.6	605	9	AV616675
17	875	13.5	563	10	BM053038
18	875	13.5	2676	11	AK014319
19	826	12.8	546	10	BF964137
20	821.5	12.7	1201	10	BF304387
21	806.5	12.5	871	12	A2682350
22	805	12.4	759	10	BF376459
23	792	12.2	662	10	BG963638
24	791.5	12.2	886	12	BH139685
25	780	12.1	517	10	BF964115
26	778.5	12.0	897	12	A2541090
27	772	11.9	780	10	BG587938
28	768	11.9	822	10	BF332761
29	766.5	11.8	932	12	A2670821
30	765	11.8	560	10	BF749379
31	764.5	11.8	834	12	A2548312
32	762.5	11.8	947	12	A2683753
33	761.5	11.8	913	12	BH155700
34	760	11.7	669	10	BM013981
35	758.5	11.7	880	12	A2687805
36	750.5	11.6	823	12	A2532602
37	748.5	11.6	853	12	A2679807
38	748.5	11.6	919	12	A2690701
39	745	11.5	699	9	BB667773
40	744.5	11.5	511	9	AA239727
41	743.5	11.5	886	12	A2540627
42	743	11.5	600	10	BM486593
43	741.5	11.5	891	12	A2682250
44	734	11.3	760	10	BM016204
45	730.5	11.3	899	12	BH154857

ALIGNMENTS

RESULT 1
AL520322
LOCUS AL520322 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB006YCL15 943 bp mRNA linear EST 13-FEB-2001
DEFINITION Prime, mRNA sequence.
ACCESSION AL520322 GI:12783815
VERSION AL520322
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 943)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..943
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB006YCL15"
/clone_lib="LTI_NFL004_NBC2"

/sex="male"

/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 253 a 225 c 253 g 211 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 4,54e-112 Length: 943
Score: 1421.00 Matches: 283
Percent Similarity: 96.41% Conservative: 12
Best Local Similarity: 92.48% Mismatches: 10
Query Match: 21.95% Indels: 1
DB: 9 Gaps: 0

US-09-672-725c-25 (1-1281) x AL520322 (1-943)

QY 977 LeuLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyClnValSerPhe 996
Db 2 CTCTTATTATTTTTCAGCTGCTGCTTTGGTGCATGGCGTGGGCAAGTCAGTTCATT 61
QY 997 AlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleGluLys 1016
Db 62 GTCCTGACTATGCCAAGCCAAATATCAGAGCCCATCATCATCATGATGAA 121
QY 1017 SerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsn 1036
Db 122 ACCCTTTGATTACAGCTACAGCAGGAAGGCTAATCCGAACACATTTGGAAGAA 181
QY 1037 ValThrPheAsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGln 1056
Db 182 GTCACATTTGGTGAAGTGTATTCACTATCCACCACCCGACATCCAGCTGCTCAG 241
QY 1057 GlyLeuSerLeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerGlyCys 1076
Db 242 GGACTGAGCTGAGGTGAAGAGGCGCACGCTGGCTCTGTTGGCGCAGTGGCTGT 301
QY 1077 GlyLysSerThrValValGlnLeuLeuGluValPheTyrAspProLeuAlaGlySerVal 1096
Db 302 GGGAAAGACACAGCTGGTCCAGCTCCTGGAGCGGTTCTAGACCCCTTGGCAGGAAAGTG 361
QY 1097 LeuIleAspGlyLysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGly 1116
Db 362 CTGCTTATCGCAAGAAATTAAGCGNCTGAATGTTTCAGTGGCTCCGAGCACACCTGGGC 421
QY 1117 IleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGly 1136
Db 422 ATCTGTGCCAGGAGCCATCTGTTTGTACTGCAGCATTTCTGAGAACATGCTGCTATGGA 481
QY 1137 AspAsnSerArgValValSerHisGluIleMetClnAlaAlaLysGluAlaAsnIle 1156
Db 482 GACAAACAGCGGGTGGTGCACAGGAGATGCTGAGGGCGCAAGAGGAGGCAACATA 541
QY 1157 HisHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThr 1176
Db 542 CATCCCTTCATCAGTCACTGCCTTAATAATATAGCACTAAAGTAGAGACAAAGAACT 601
QY 1177 GlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaIleAlaValArgGlnPro 1196
Db 602 CAGCTCTCTGTTGGCCAGAAACACGCAATTCGCATAGCTCTGCTCCCTTGTAGACAGCCT 661
QY 1197 HisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValVal 1216
Db 662 CATATTTGCTTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGGTTGTC 721

QY 1217 GlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeu 1236
Db 722 CAAGAACCCTGGACAAAGCCAGAGAGCGCCACCTGCTGATGTGCTACCGCCTG 781
QY 1237 SerThrIleGlnAsnAlaAspLeuIleVal-ValPheGlnAsnGlyLysValLysGluHi 1256
Db 782 TCCACCATTCCAGATGTCAGACTTAATAGTGTGTGTTTTCAGATGGCAGAGCA 841
QY 1256 sGlyThrHisGlnGlnLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGI 1276
Db 842 TGGCACCATCAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGTCAGTGTC 901
QY 1276 nAlaGlyAlaLysArg 1281
Db 902 GCCTGGACAAAGCGC 917
RESULT 2
BF969667 981 bp mRNA linear EST 22-JAN-2001
LOCUS 602272046F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4360090 5',
DEFINITION mRNA sequence.
ACCESSION BF969667
VERSION BF969667.1 GI:12336882
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 981)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10000 Row: 1 Column: 11
High quality sequence stop: 670.
FEATURES
source
1. 981
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4360090"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site: 1:
NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 267 a 194 c 250 g 270 t
ORIGIN
Alignment Scores:
Pred. No.: 3,67e-82 Length: 981
Score: 1073.50 Matches: 234
Percent Similarity: 86.60% Conservative: 18
Best Local Similarity: 80.41% Mismatches: 31
Query Match: 16.58% Indels: 10
DB: 10 Gaps: 5
US-09-672-725c-25 (1-1281) x BF969667 (1-981)
QY 793 MetValPheArgSerMetLeuArgGlnAspValSerTrpPheAspProLysAsnThr 812
Db 1 ATGGTTTCCGATCCATGCTCAGACAGGATGTGAGTTGGTTGATGACCCATAAACACC 60

QY 813 ThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaAlaGlnValLysGlyValaIleGly 832
 Db 61 ACTGGAGCATGACTACAGGCTGCCAATGATGCTGCTCAAGTAAAGGGCGCTATAGGT 120
 QY 833 SerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleSer 852
 Db 121 TCCAGGCTGCTGTAATACCCAGAAATATAGCAATCTTGGACAGGAATAATATATCC 180
 QY 853 LeuIleTyrGlyTyrGlnLeuThrLeuLeuLeuAlaIleValProIleIleAlaIle 872
 Db 181 TTCATCATGTTGGCACTAACACTGTTACTCTAGCAATGTACCCATCATGTGCAATA 240
 QY 873 AlaGlyValValGluMetLysMetLysSerGlyGlnAlaLeuLysAspLysLysGluLeu 892
 Db 241 GCAGGAGTGTGAATGAAATGAAATGTTCTGGCAACGACTGAAAGATAAGAAAGACTA 300
 QY 893 GluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeu 912
 Db 301 GAAGTGCTGGGAAGATCGCTACTGAAGCAATAGAAACTTCCGAACCGTTGTCTTTG 360
 QY 913 ThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsn 932
 Db 361 ACTCAGCAGCAGAGTTGACATATGATGCTCAGAGTTTGAGGATACCATACAGAAAC 420
 QY 933 SerLeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyr 952
 Db 421 TCTTTGAGGAAGCACACATCTTTTGAATACATTTTCTTCCACCCAGGCAATGATGAT 480
 QY 953 PheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsn 972
 Db 481 TTTTCTTATCTGGATGTTTCCGGTTTGGAGCTTACTTGTGTGCACATAAATCATGAGC 540
 QY 973 PheGlnAspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGln 992
 Db 541 TTTGAGGATGTTCTGTAGTATTTTTCAGCTGTTGCTTGTGGTCCATGGCGGGGCAA 600
 QY 993 ValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMet 1012
 Db 601 GTCAGTTCAATTTGCTCTGACTATGC-AAAGCCAAATATATCAGCAGC-CACATCATG 658
 QY 1013 IleIleGluLysSerProLeuIle---AspSerTyrSerProHisGlyLeuLysProAsn 1031
 Db 659 ATCATGAAAAACCCCTGTGTTGATTTGACAGTTGACAGCGGGAAGGCTAATTTGGCCGA 718
 QY 1032 ThrLeu---GluGlyAsnVal---ThrPheAsnGluValValPheAsnTyrProThrArg 1049
 Db 719 ACATTTGGGAAGGAATGCTCTTCTTGGTGAGGTTGGATTTCATGATCCCAACCGA 778
 QY 1050 ProAspIleProValLeuGlnGly-----LeuSerLeuGluValLysLysGlyGlnThr 1067
 Db 779 CCGGAATTCCTCAGTGTCAAGGATGAGCCCGGGGTAAAAAGGGCCAAACGCGTCTCTTT 838
 QY 1068 LeuAlaLeuValGly-----SerSerGly 1075
 Db 839 GGAGCATGTGTGGGAACCATGTCCTCTGGG 871

RESULT 3
 BF584668
 LOCUS 602098406F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4218385 5',
 DEFINITION mRNA sequence.
 ACCESSION BF584668
 VERSION BF584668.1 GI:11658386
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 894)
 NIH-MGC <http://mgc.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAM9798 row: 1 column: 02
 High quality sequence stop: 651.

FEATURES

Location/Qualifiers

source

1..894
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:4218385"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="organ: colon; Vector: pCMV-SPORT6; Site_1: Noti;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.6 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 263 a 203 c 242 g 185 t
 ORIGIN

Alignment Scores:

Pred. No.: 6,72e-79 Length: 894
 Score: 1035.00 Matches: 225
 Percent Similarity: 92.48% Conservative: 21
 Best Local Similarity: 84.59% Mismatches: 15
 Query Match: 15.99% Indels: 9
 DB: 10 Gaps: 1

US-09-672-725c-25 (1-1281) x BF584668 (1-894)

QY 335 ThrValPheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIle 354
 Db 2 ACTGTCTCTTTCGCTGTTAAT-GGAGCATTCAGTGTGGACAGCATCTCCAATATT 60
 QY 355 GluAlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLys 374
 Db 61 GAAGCCTTCGCCAATGCACGAGGAGCAGCTTATGAAGTCTTCAAAATAATTGATAATAAG 120
 QY 375 ProSerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeu 394
 Db 121 CCCAGTATAGACAGCTTCTCAAAAGAGTGGGCACAAACAGACACATACAGGAATCTG 180
 QY 395 GluPheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeuLysGly 414
 Db 181 GAATTTAAGAAATATTCACCTTCAGTTACCCATCTCGAAAGAAGTTCAGATCTTGAAGGCG 240
 QY 415 LeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGly 434
 Db 241 CTCAATCTGAAGGTGAAGAGCGGACAGACGGTGGCCCTGGTTGGCAACAGTGGCTGGA 300
 QY 435 LysSerThrThrValGlnLeuMetClnArgLeuTyrAspProThrAspGlyMetValCys 454
 Db 301 AAAACACAACTGTCCAGCTGATGCAAGGCTCTACGACCCCTAGATGGCATGGTCAGT 360
 QY 455 IleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThrGlyVal 474
 Db 361 ATCAGCGCACAGGACATCAGAACCATCAATGTGAGGTATCTGAGGGAGATCATTTGGTGTG 420
 QY 475 ValSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArg 494
 Db 421 GTGAGTCAGGAACCTGTCTGTTGCCACACGATCGCCGAGAACATTCGCTATGGCGGA 480
 QY 495 GluAsnValThrMetAspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 514
 Db 481 GAAGATGTCAACCATGGATGAGATTGAGAAAGCTGTCAAGGAAGCAATGCTATGACTTC 540
 QY 515 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSer 534

```
|||||
Db 541 ATCATGAAGTCCCAACAAATTGACACCTGGTGGTGAGAGGGCGCAGCTGAGT 600
QY 535 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu 554
|||||
Db 601 GGGGGACAGAAACAGAGAATCGC-ATTGGCCGGGCCCTGGTCCGCAATCCCAAGTGCCTT 659
QY 555 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValGlnValAla 574
|||||
Db 660 TC-GTGACGAGACCCACCTCAGC-CTGGATACAGAAAGTGAAGCTGGGGTTCCAGGCCGCA 717
QY 575 -LeuAspLysAlaArgLysGlyArgThrThrIle-----ValIleAlaHisArg 590
|||||
Db 718 CCTAGATAGGTACAAAGGCGNGGACCAATGGGAAGCTCATCGCTGTACCGTCGAA 777
QY 590 gLeuSerThrValArg 595
|||||
Db 778 GCTGACGGCCATCGCG 793

RESULT 4
BG298756 803 bp mRNA linear EST 21-FEB-2001
LOCUS 602396681F1 NTH_MGC_94 Mus musculus cDNA clone IMAGE:4511157 5',
DEFINITION mRNA sequence.
ACCESSION BG298756
VERSION BG298756.1 GI:13063728
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 803)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM10394 row: b column: 22
High quality sequence stop: 698.

FEATURES
source
Location/Qualifiers
1..803
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4511157"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 240 a 162 c 205 g 196 t
ORIGIN

Alignment Scores:
Pred. No.: 7.2e-79 Length: 803
Score: 1034.00 Matches: 223
Percent Similarity: 91.51% Conservative: 14
Best Local Similarity: 86.10% Mismatches: 21
Query Match: 15.97% Indels: 4
DB: 10 Gaps: 0

US-09-672-725C-25 (1-1281) x BG298756 (1-803)
QY 236 LysIleLeuSerSerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaVal 255
```

```
|||||
Db 3 AAAGATATGCTTTCATTACTGATAAGAACTCCATGCTTATGCAAAAGCTGGAGCAGTT 62
QY 256 AlaGluGluValLeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGlu 275
|||||
Db 63 GCTGAAGAAGTCTTAGCAGCCATCAGAACTGTGATTCGGTT-GGAGGACAAAAGAGGAA 121
QY 276 LeuGluArgTyrAsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIle 295
|||||
Db 122 CTTGAAAGGTACAATAACAACCTTGGAAAGCTAAAAGGCTGGGGATAAAAAGACGTATC 181
QY 296 ThrAlaAsnIleSerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAla 315
|||||
Db 182 ACGGCCAACATCTCCATGGGTGCAGCTTTCTCCTTATCTATGCTCATATGCTCTGGCA 241
QY 316 PheTyrGlyThrSerLeuValLeuSerSerGluTyrSerIleGlyGlnValLeuThr 335
|||||
Db 242 TTCGTGATGGGACTTCTCTGGTCATCTCCAAAGAACTCTAT-GGACAAGTGCTCACT 300
QY 336 ValPhePheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGlu 355
|||||
Db 301 GTCTTCTTTTCGGTGTAAATGGAGCATTCAGTGTGGACAGGCATCTCCAAATATTGAA 360
QY 356 AlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysPro 375
|||||
Db 361 GCCTTCGCCAATGACGAGGAGCAGCTTATGAAGTCTTCAAAATATTTGATAAAGCCC 420
QY 376 SerIleAspSerTyrSerLysSerGlyHisLysProAsnIleLysGlyAsnLeuGlu 395
|||||
Db 421 AGTATAGACAGCTTCTCAAGAGTGGGCACAAACCCAGACACATACAAAGGAATCTGGAA 480
QY 396 PheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeu 415
|||||
Db 481 TTTAAGAATATTCATCTCAGTTACCCATCTCGAAAAGAGTTCAGATCTTGAAGGGCCTC 540
QY 416 AsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLys 435
|||||
Db 541 AATCTGAAGTGAAGAGCGGACAGCGTGGCCCTGGTTGGCAACAGTGGCTGTGAAAA 600
QY 436 SerThrThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIle 455
|||||
Db 601 AGCAACAACCTGTCAGCTGATGCAAAAGGCTCTACGACCCCTA-GATGGCATGTCAGTATC 659
QY 456 AspGlyGlnAspIleArgThrIle-AsnValArgHisLeuArgGluIleThrGlyValVa 475
|||||
Db 660 GACGGCAGGACATCAGAACCATCAATGTGAGGTATCTGGAGGGGATCATGGGGTGTG 719
QY 475 lSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGly 493
|||||
Db 720 AGTCAAGGAACCTGGCTTGTGTTGCCAACCCAGATCCGAGAAAATTCGTTATGGG 774

RESULT 5
BG248052 1019 bp mRNA linear EST 13-FEB-2001
LOCUS 602359987F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4488404 5',
DEFINITION mRNA sequence.
ACCESSION BG248052
VERSION BG248052.1 GI:12757867
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 1019)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: L1AM10334 row: n column: 21
High quality sequence stop: 650.
Location/Qualifiers
1. 1019

FEATURES

source
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:448404"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: Sali; Site.2: Noti; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT 276 a 262 c 283 g 196 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 6.94e-78 Length: 1019
Score: 1024.00 Matches: 215
Percent Similarity: 90.48% Conservative: 12
Best Local Similarity: 85.71% Mismatches: 19
Query Match: 15.82% Indels: 5
Db: 10 Gaps: 2

US-09-672-725c-25 (1-1281) x BG248052 (1-1019)

QY	1033	LeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIle	1052
Db	1	TGGGAAGAAATGCAATTTAGTGAGTCTGTCAACTATATCCACCGACCCGACATC	60
QY	1053	ProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGly	1072
Db	61	CCAGTCTTCAGGGCTGAGCTTGAGTGAAGAGGGCCAGAGCGTGGCCGTGGGC	120
QY	1073	SerSerGlyCysGlyLysSerThrValValcInLeuLeuGluArgPheTyrAspProLeu	1092
Db	121	AGCAGTGGCTCGGGAAGAGACACAGTGGTCCAGTCTCGAGCGCTTCTACGACCCCATG	180
QY	1093	AlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArg	1112
Db	181	GCTGGATCAGGTTCCTAGATGGCAAGAAATAAAGCAACTGAATGTCCAGTGGCTCCGA	240
QY	1113	AlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsn	1132
Db	241	GCACAGTGGGCTGTGTCCACAGAGCCATCTCTTTGACATGCACATCGCAGAGAAC	300
QY	1133	IleAlaTyrGlyAspAsnSerArgValValSerHisGluLeuMetGlnAlaAlaLys	1152
Db	301	ATTGCTACGGAGACAACAGCGGGTCTGTCTTATCAGGAGATTGTGAGGGCAGCCAAAG	360
QY	1153	GluAlaAsnIleHisHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGly	1172
Db	361	GAGGCCAATCCACCATGTTTCATCGATCGCTACCTGATAAATACACACAGAGTAGGA	420
QY	1173	AspLysGlyThrGlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeu	1192
Db	421	GACAAAGGCACCTACGTGTGGGTGGGCAGACAGCGCATCGCCATCGCAGCGCCCTC	480
QY	1193	ValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer	1212
Db	481	GTCAGACAGGCTCACATTTTACTTCTGCGAAGACACATCAGCTCTGGATACAGAAAGT	540
QY	1213	GluLysValValGlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIle	1232
Db	541	GAAGAAGTTGTCCAGGAAGCGCTGGACAAAGCCAGGGGAAGCGCACCTCATTTGTATC	600
QY	1233	AlaHisArgLeuSerThrIleGlnAsnAlaAspLeuIle-ValVal-PheGlnAsnGlyL	1252

Db	601	GCTCACGGCTGTCCACCATCCAGAGACGGAGCTTGTATCGTGTGATTCACAGACCGGA	660
QY	1252	ysValLys---GluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPhes	1271
Db	661	AGGTCAAGGAGGACGAGGGGACCCCAACAGCAAGTGGTGGCGCAGAGGCGCATCTACTTCT	720
QY	1271	erMetValSer-ValGlnAlaGlyAlaLys	1280
Db	721	CCAGG---TCAGGTGCATCTGGCGCAAG	747
RESULT	6		
LOCUS	BF796582	944 bp mRNA linear EST 12-JAN-2001	
DEFINITION	602258463F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4341710 5',		
ACCESSION	BF796582		
VERSION	BF796582.1 GI:12101636		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 944)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue procurement: Louis Staudt, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1AM9555 row: n column: 15 High quality sequence start: 9 High quality sequence stop: 669. Location/Qualifiers 1. 944 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4341710" /clone_lib="NIH_MGC_85" /tissue_type="lymphoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: lymph; Vector: pCMV-SPORT6; Site.1: Noti; Site.2: Sali; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."		
BASE COUNT	265 a 241 c 237 g 201 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	3.76e-73 Length: 944		
Score:	968.50 Matches: 210		
Percent Similarity:	83.64% Conservative: 20		
Best Local Similarity:	76.36% Mismatches: 39		
Query Match:	14.96% Indels: 8		
Db:	10 Gaps: 1		
US-09-672-725c-25 (1-1281) x BF796582 (1-944)			
QY	988	MetAlaValGlyGlnValSerPheAlaPro-Asp-TyrAlaLysAlaLysValSerA	1007
Db	1	GTGGCTCTAGGACATCCAGTTCATTTGCTCCAGACGATGCTAAAGCTAAGCTGTCTG	60
QY	1007	laAlaHisValIleMetIleGluLysSerProLeuIleAspSerTyrSerProHisG	1027
Db	61	CAGCCCACTTATTCATGCTGTTTGAAGACAACTCTGATTGACAGCTACAGTGAAGAGG	120

```
Qy 1027 lyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrP 1047
Db 121 GCGTGAAGCCTGATAAATTTGAAGAAATATAACATTTAATGAAGTCGTGTTCAACTATC 180
Qy 1047 roThrArgProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnT 1067
Db 181 CCACCGGACCAACGTCGACGTTCAGGGCGTGGAGGTGAAGAAAGGCCAGA 240
Qy 1067 hrLeuAlaLeuValGlySerGlyCysGlyLysSerThrValValGlnLeuLeuGluA 1087
Db 241 CACTAGCCCTGGTGGCAGCATGCTGTGGGAAGACAGCGGTGTCTCAGCTCTCTGGAGC 300
Qy 1087 rgPheTyrAspProLeuAlaGlySerValLeuIleAspGlyLysGluLysHisLeuA 1107
Db 301 GPTTCTACGACCCCTGGCGGGACAGTCTCTCGATGGTCAAGAAAGCAAAACTCA 360
Qy 1107 snValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspC 1127
Db 361 ATGTCCAGTGGCTCAGAGCTCAACTCGGAATCGTGTCTCAGAGCGCTATCCTATTGTGACT 420
Qy 1127 ysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgValValSerHisGluGluI 1147
Db 421 GCAGCATTCGCCGAGATATGCTGTATGGAGACAAACAGCGGGTGTGTATCAGAGATGAA 480
Qy 1147 leMetGlnAlaLysGluAlaAsnIleHisPheIleGluThrLeuProGluLysT 1167
Db 481 TCGTGAGTCAGCAAAAGCTGCCAACATACATCCTTTTCAGAGACGTTACCCCAAAAT 540
Qy 1167 yrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGlnLysGlnArgIleA 1187
Db 541 ATGAACAAGAGTGGGAGATAGGGGACTCAGCTCTCAGAGGGTGCACAAACAGAGATTG 600
Qy 1187 laIleAlaArgAlaLeuValArgGln-ProHisIleLeuLeuLeuAspGluAlaThrSer 1206
Db 601 CTATTGCCGCGCCCTCATCAGACACCTCAAACTCTCTCTGTGGATGAGCTACATCA 660
Qy 1207 Ala-LeuAspThr-----GluSerGluLysValValGlnGluAlaLeuAspLysAlaAr 1224
Db 661 GCCTCTGGATACCTGCAAGTAGTGAACAGCGGTGTGCCAAGAAGCCCTGTGCACACAGCAGA 720
Qy 1224 gGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAlaAspLe 1244
Db 721 AGAGCCGCGCACCTGGCTGTGTGATTGCTCACCCTGCTC--ACATCCAGAAATGCAGACTT 778
Qy 1244 uIleValValPheGlnAsnGlyLysValLysGluHis 1256
Db 779 -ATAGTGGTGTCCCAACGGGAGAGTCAAGACACAT 814

RESULT 7
BF969062 851 bp mRNA linear EST 22-JAN-2001
LOCUS 602269993F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4358135 5',
DEFINITION mRNA sequence.
ACCESSION BF969062
VERSION BF969062.1 GI:12336277
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 851)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

```
Plate: LAM9995 row: j column: 24
High quality sequence stop: 612.
Location/Qualifiers
source 1..851
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4358135"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 243 a 172 c 214 g 222 t
ORIGIN
Alignment Scores: 2.65e-72 Length: 851
Pred. No.: 958.00 Matches: 217
Score: 79.25% Conservative: 16
Percent Similarity: 73.81% Mismatches: 33
Best Local Similarity: 14.80% Indels: 29
Query Match: 10 Gaps: 5
DB:

US-09-672-725C-25 (1-1281) x BF969062 (1-851)

Qy 826 GlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeu 845
Db 3 CAAGTTAAAGGGGTATAGGTTCCAGGCTTGCTGTAATACCCAGAAATATAGCAAAATCTT 62
Qy 846 GlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAla 865
Db 63 GGGCAGAGAAATATATATCTTCATCTATGTTGGCAACTTAACACTGTACTCTTAGCA 122
Qy 866 IleValProIleIleAlaIleAlaGlyValValGluMetLysMetLysSerGlyGlnAla 885
Db 123 ATTGTACCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAAATGTTGCTCGGACAAGCA 182
Qy 886 LeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsn 905
Db 183 CTGAAAGTAAAGAAGAACTAGAGAGTGTCTGGGAAGATCGCTACTGAAGCAATAGAAAAC 242
Qy 906 PheArgThrValValSerLeuThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSer 925
Db 243 TTCGGAACCGTGTCTTCTTGACTCAGGAGCAGAAGTTTGAACATATATGTATGCTCAGAGT 302
Qy 926 LeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGlyValSerPheSer 945
Db 303 TTGCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTTTGGAATTACATTTTCC 362
Qy 946 IleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeu 965
Db 363 TTCACCCAGGCAATGATGATATTTTTCCTATGCTGATGTTTCCGGTTTGGAGCCTACTTG 422
Qy 966 ValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSerAlaIleValPhe 985
Db 423 GTGGCACAATAAACAATCATGAGCTTTGAGGATGTTCTGTAGTATTTTCAGCTGTGTCTTT 482
Qy 986 GlyAlaMet-AlaValGlyGlnValSerSerPheAlaProAspTyrAlaLysAlaLysVa 1005
Db 483 GTGGCCATGGGCCGTGGGCAAGTCAGTTTCATTGCTCTGACTATGCCAAGCAAAAT 542
Qy 1005 lserAlaAlaHisValIleMetIleIleGluLysSerProLeuIleAspSerTyrSerPr 1025
Db 543 ATCAGCAGCCCACTCATCATGATCATTTGAAATAAACCCCTTTGAT-GACAGCTACAGCAC 601
Qy 1025 ohisGlyLeuLysProAsnThr---LeuGluGlyAsnValThrPhe-----AsnGluVa 1042
Db 602 GGAAGGCCCTAAATTGCCGGAACACATTTGGGAAGGAAATGTCCACCATGTGGTGGAACGTTGT 661
Qy 1042 lvalPheAsnTyrProThrArg-ProAspIleProValLeuGlnGlyLeuSerLeuGluV 1062
```



```

Db 662 TTTTCAATTATTCACACCGTAGCGGACATCCAGCGTCTCCAGGACTGAAGCTGGGGT 721
      |||||
QY 1062 alLysLysGlyGlnThrLeuAlaLeu-----ValGlySerSer----- 1074
      :|||
Db 722 GAAGAAAGGGCGGCGGCTCTCGGGTGGGACCAATGGTGGTGGGAAAGACACCA 781
      |||||
QY 1075 -----GlyCysGlySerThrValValGlnLeuLeuGluArgPheTyrA 1090
      |||
Db 782 GGGGGCCAGTCCCGGACCGGTCTAAACC----- 813
      |||||
QY 1090 spProLeuAlaGlySerValLeuLeuAspGlyLysGlu 1102
      |||||
Db 814 -----CTGGGGGAAATGCTGCTCTGGGCAAAAAAAG 846
      |||||

RESULT 8
AV709991
LOCUS AV709991 785 bp mRNA linear EST 09-OCT-2000
DEFINITION AV709991 Cu Homo sapiens cDNA clone CUAJA09 5', mRNA sequence.
ACCESSION AV709991
VERSION AV709991.1 GI:10728272
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 785)
AUTHORS Peng, Y., Song, H., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N.,
Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu
W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J.,
Chen, Z. and Han, Z.
Homo sapiens cDNA Clones
Unpublished (2000)
CONTACT: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
location/Qualifiers
    1..785
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="CUAJA09"
    /clone_lib="Cu"
    /tissue_type="adrenal cortico adenoma for Cushing's
    syndrome"
    /dev_stage="Adult"
    /note="Vector: pBluescript sk(-)"
BASE COUNT 210 a 180 c 211 g 183 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 1..18e-71 Length: 785
Score: 950.00 Matches: 208
Percent Similarity: 86.87% Conservative: 17
Best Local Similarity: 80.31% Mismatches: 33
Query Match: 14.68% Indels: 4
DB: 9 Gaps: 0

US-09-672-725C-25 (1-1281) x AV709991 (1-785)

QY 955 TyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGln 974
      |||
Db 13 TACTAAGGCTGCAGCCCTTGGCAATTCAGAGTATGTTTACTTTCCTTGATTAAGAGA 72
      |||
QY 975 AspValLeuLeuPheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSer 994
      |||
Db 73 GGAATTGGAAGAGATATTTCACCTGTTGCTTTGGTGGCCAGTGGCGGCAAGTCACT 132
      |||
QY 995 SerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMetIle 1014
      |||

```

```

Db 133 TCATTTCCTCCTGACTATGCCAAGCCAAATATATCAGCAGCCACATCATGATCATTT 192
      |||||
QY 1015 GluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGlu 1034
      |||||
Db 193 GAAAAAACCCCTTGTATTGACAGCTACACACGGAAGGCTTAATGCCGAACACATTGGAA 252
      |||||
QY 1035 GlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIleProVal 1054
      |||||
Db 253 GGAATATGTACATTTTGGTGAAGTTGATTCAACTATCCACCCGACCGACATCCAGTG 312
      |||||
QY 1055 LeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSer 1074
      |||||
Db 313 CTTCAGGACTGAGCCTGGAGGTGAAGAGGGCCAGACGCTGGCTCTGGTGGGACAGCT 372
      |||||
QY 1075 GlyCysGlyLysSerThrValValGlnLeuLeuGluAArgPheTyrAspProLeuAlaGly 1094
      |||||
Db 373 GGCTGTGGGAAGACACACAGCTGTCCAGCTCTCTGGAGCGGTCTACGACCCCTTGGCAGG 432
      |||||
QY 1095 SerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHis 1114
      |||||
Db 433 AAAGTCTGCTGTATGGCAAGAAATTAAGCGACTCAATGTTTCAGTGGCTCCGAGCACAC 492
      |||||
QY 1115 LeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAla 1134
      |||||
Db 493 CTGGGATCGTGTCCAGGAGCCCATCTCTTTGACTCGAGCATTTGTCAGAACATTTGCC 552
      |||||
QY 1135 TyrGlyAspAsnSerArg-ValValSerHisGluGluIleMetGlnAlaAlaLysGluAl 1154
      |||||
Db 553 TATGGAGACAACAGCCGGGTGTGTCTCAGAGAACAGATTTGAGGGCAGCAAGAGGAGGC 612
      |||||
QY 1154 asnIleHisHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAspIly 1174
      |||||
Db 613 CAACATACATGCTTCATCGAGTCACTGCCTAT-AAATATAGCAGCTAAACTANGAGACAA 671
      |||||
QY 1174 sglyThrGlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValAr 1194
      |||||
Db 672 AGGACCTCAGCTCTCTGGGGCCAGAAACA-CACATTGCCATAGCTTGTGCCCTGGT-AG 729
      |||||
QY 1194 gGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer 1212
      |||||
Db 730 ACAGACTTATAATTGTTTGGATGAAGCCAGCTCAGCTCTTGATACAGAAAGT 784
      |||||

RESULT 9
BF313560 795 bp mRNA linear EST 21-NOV-2000
LOCUS 601900192F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:412922 5',
      mRNA sequence.
ACCESSION BF313560
VERSION BF313560.1 GI:11261583
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 795)
AUTHORS NTH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@rsr@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI023 row: h column: 23
High quality sequence stop: 674.
location/Qualifiers
    1..795
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
FEATURES
source

```

```

/clone="IMAGE:4129222"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: brain; Vector: pOTB7; site_1: XhoI; site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 213 a 189 c 236 g 157 t
ORIGIN

```

Alignment Scores:
 Pred. No.: 9.08e-69 Length: 795
 Score: 916.50 Matches: 210
 Percent Similarity: 87.50% Conservativeness: 14
 Best Local Similarity: 82.03% Mismatches: 29
 Query Match: 14.16% Indels: 9
 DB: 10 Gaps: 2

US-09-672-725C-25 (1-1281) x BF313560 (1-795)

```

QY 1017 SerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsn 1036
DB 2 ACCCCCTTTGATGACAGCTACAGCAGCGGAGCCCTAATGCCGACACATG-GAAGGAAT 60

QY 1037 ValThrPheAsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGln 1056
DB 61 GTCACATT-GGTGAAGT-GTATTCAACTATCCACCGACCGACATCCAGTGTTCAG 118

QY 1057 GlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSerglyCys 1076
DB 119 GGACTGAGCTGGAGTGGAAGGCGCAGCGTCTGGTGGGAGCAGTGGCTGT 178

QY 1077 GlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerVal 1096
DB 179 GGAAGAGCAGACAGTGTCTCCAGCTCTGGAGCGGTTCTAGCACCCCTTGGCAGGGAAGTG 238

QY 1097 LeuIleAspGlyLysGluIleLysHisLeuAsnValGlnThrPheArgAlaHisLeuGly 1116
DB 239 CTGCTTGTATGGCAAGAAATAAAGCGACTGAATGTTCAGTGGCTCCGAGCACACCTGGGC 298

QY 1117 IleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGly 1136
DB 299 ATCGTGTCCAGGAGCCCATCTCTTTGACTGCAGCATTTGCTGAGACATTCCTATGGA 358

QY 1137 AspAsnSerArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIle 1156
DB 359 GACAACAGCGGGTGTGTACAGGAAGAGATCGTGAGGGCAGCAAGAGGAGGCCAACATA 418

QY 1157 HisHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThr 1176
DB 419 CATGCCCTTCATCGAGTCACTGCTGCTTAATAATATAGCATTAAGTAGGACAAAGAACT 478

QY 1177 GlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnPro 1196
DB 479 CAGCTCTCTGGTGGCCAGAAACAAACGCATTCGATAGCTCGTGGCCCTTTAGACAGCT 538

QY 1197 HisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluIysVal-Va 1216
DB 539 CATATTTT-GCTTTGGTGAAGCCAGCTCAGCTCTGGATACAGAAAGTAAAAAGTGCCT 596

QY 1216 lGlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLe 1236
DB 597 CCAAGAAAGCCCTGGACAAAGCCAGAGAAGGC---GACCTGCATGTGTGTGTACCGCTG 653

QY 1236 uSerThrIleGlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHi 1256
DB 654, -TCCACCATCCAGATGAGA -CTTATAGTGGTGTTCAGAAATGGCAGAGTCCAGGGGC 710

```

```

QY 1256 sGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSer 1271
DB 711 ATGGAGCATCAGAGTGTGGACAGAAAGG---CTCTATTTTCA 753

RESULT 10
BG293345 726 bp mRNA linear EST 21-FEB-2001
LOCUS 602390738F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4502552 5',
DEFINITION mRNA sequence.
ACCESSION BG293345
VERSION BG293345.1 GI:13052943
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 726)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10371 row: 1 column: 09
High quality sequence stop: 658.
Location/Qualifiers
1..726
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4502552"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 189 a 220 c 190 g 127 t
ORIGIN

Alignment Scores:
Pred. No.: 8.99e-69 Length: 726
Score: 916.00 Matches: 185
Percent Similarity: 94.66% Conservativeness: 10
Best Local Similarity: 89.81% Mismatches: 9
Query Match: 14.15% Indels: 2
DB: 10 Gaps: 0

US-09-672-725C-25 (1-1281) x BG293345 (1-726)
QY 1068 LeuAlaLeuValGlySerSerglyCysGlyLysSerThrValValGlnLeuLeuGluArg 1087
DB 2 CTGGCCCTGTGGTGGCAGCAGTGGCTGGGGAGAGACAGTGTCCAGCTGCTCGAGCGC 61

QY 1088 PheTyrAspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsn 1107
DB 62 TTCATGACCCCATCGCTGGATCAGTGTCTTAGTGTGTCAGGAAGCAAGAACTCAAT 121

QY 1108 ValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCys 1127
DB 122 GTCCAGTGGCTCCGAGCTCAACTGGGCATTGTGCCAGGAACCACTCTCTTTGACTGC 181

QY 1128 SerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgValSerHisGluGluIle 1147
DB 182 AGCATCGAGAGAACATCGCCTATGGAGACACAGCCGGTGGTCCCTCATGATGATGAGATT 241

```

QY 1148 MetGlnAlaLysGluAlaAsnIleHisPheIleGluThrLeuProGluLysTyr 1167
 Db 242 GTGAGGGCAGCCAGGAGCCACATCCCTTCATCGAGAGCGTCCCAAAATAT 301
 QY 1168 AsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleAla 1187
 Db 302 AACACAGAGTAGGAGACAGGGAGCGAGCTCTCTGGGGGCCAGAGAGGATGGCC 361
 QY 1188 IleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAla 1207
 Db 362 ATCGCCGAGCCCTCATCAGACAGCTGGCTCTACTGCTGGATGAGCCACGTCAGCT 421
 QY 1208 LeuAspThrGluSerGluLysValValGlnGluAlaLeuAspLysAlaArgGluLysArg 1227
 Db 422 CTGGATCTAGAGTGAAGGTTGTCCAGGAAGCATGTGCACAAAGCCAGGAAGGCCGC 481
 QY 1228 ThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAla-AspLeuIleValVa 1247
 Db 482 ACCTGCATTGCGATCGCTACCGCTGTCCACCATCCAGAACGGCGCACTTGATCGTGGT 541
 QY 1247 lPheGlnAsnGlyLysValLysGluHisGlyThr-HisGlnGlnLeuLeuAlaGlnLysG 1267
 Db 542 GATTGAGAACGGCAAGGTCAAGGACGACGGCCACCCACGACGCTGCTGGCCGAGAAGG 601
 QY 1267 lylleTyrPheSer 1271
 Db 602 GCATCTATTCTCA 615
 RESULT 11
 BM053000
 LOCUS
 DEFINITION
 ie69a07.y3 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 cDNA 5' similar to SW:MDR1_HUMAN P08183 MULTIDRUG RESISTANCE
 PROTEIN 1 ;, mRNA sequence.
 ACCSSION
 VERSION
 KEYWORDS
 SOURCE
 BM053000.1 GI:1680894
 EST.
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemishka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blustain, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
 , M., Gibbons, M., McCann, R., Cole, R., Tsagarelshvili, R., Williams, T.,
 Jackson, X. and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brown@fas.harvard.edu)
 High quality sequence stop: 437.
 Location/Qualifiers
 i. .587
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
 /sex="Both"
 /tissue_type="Islets of Langerhans"
 /dev_stage="Adult"
 /lab_host="DH10B"
 /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
 FEATURES
 source

Site_2: Sal 1; Starting library constructed using
 SuperScript Plasmid Library kit (Life Technologies). cDNA
 made by oligo-dT priming. Size-selected by column
 fractionation; average insert size 1.08 kb. Library was
 amplified once on solid support and plasmid DNA from
 library was prepared. The library DNA was normalized by
 method #4 from Bonaldo, Lennon, and Soares 1996 Genome
 Research 6:791-806; 0.5 microgram single-stranded library
 plasmid DNA was mixed with 5 micrograms PCR product
 representing library inserts and hybridized to an Ecot of
 20. Single-stranded (unhybridized) plasmids were isolated
 by hydroxyapatite chromatography and used to make this
 library."

BASE COUNT 142 a 138 c 161 g 145 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 2,29e-68 Length: 587
 Score: 910.00 Matches: 175
 Percent Similarity: 96.34% Conservative: 9
 Best Local Similarity: 91.62% Mismatches: 7
 Query Match: 14.06% Indels: 0
 DB: 10 Gaps: 0

US-09-672-725C-25 (1-1281) x BM053000 (1-587)

QY 950 MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrIleuValAlaAsnGlu 969
 Db 3 ATGATGTATTTTCCCTATGCTGATGTTTCCGGTTTGGAGCCTACTTGTGTGCACATAAA 62

QY 970 PheMetAsnPheGlnAspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAla 989
 Db 63 CTCATGAGTTTTCAGGATGTTCTGTAGTATTTTTCAGCTGTGTCTTTTGTGTGCCATGCC 122

QY 990 ValGlyGlnValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaHis 1009
 Db 123 GTGGGCGAGTCAGTTCATTTGCTGCTGACTATGCCAAGCCAAATATACAGACGCCAC 182

QY 1010 ValIleMetIleIleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys 1029
 Db 183 ATCATCATGATCATTTGAAAAAACCCTTTTGATTGACAGCTACAGCAGGAGGCTTAATG 242

QY 1030 ProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArg 1049
 Db 243 CCGAACACATTGGAAGGAATCTCACATTTGCTGAAGTTGTATTCAACTATATCCACCGCA 302

QY 1050 ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrLeuAla 1069
 Db 303 CCGGACATCCAGTCTTCAGGGAGTCTGAGCTGGAGGTGAAGAGGCGCAGCGTGGCT 362

QY 1070 LeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyr 1089
 Db 363 CTGGTGGGCGAGTGGCTGTGGGAAGAGACAGTGGTCCAGCTCCTGGAGCGGTTCAC 422

QY 1090 AspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGln 1109
 Db 423 GACCCCTTGGCAGGAAAGTGTCTGTTGCAAAAGAAATAAAGCGCATGAATGTTTCAG 482

QY 1110 TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle 1129
 Db 483 TGGCTCCGAGCACACCTGGGCATCGTGTCCAGGAGGCCATCTCTGTTGACTGTCAGCAT 542

QY 1130 AlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
 Db 543 GCTGGAACATGCTCTATGAGACAAACAGCCGN 575

RESULT 12
 BG080311
 LOCUS
 DEFINITION
 H3052B06-5 NTA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 H3052B06 5', mRNA sequence.
 BG080311
 ACCESSION
 BG080311.1 GI:12562879

628 bp mRNA linear EST 26-JAN-2001

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 628)
Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka
T.S., Carter, M.G. and Ko, M.S.H.
TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: H3052B06-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 400, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit <http://igsun.grc.nia.nih.gov/cDNA/15K.html> for details.
Plate: H3052 row: B column: 06
Seq primer: -21M13 Reverse
High quality sequence stop: 628
POLYA-No.

FEATURES
source Location/Qualifiers
1..628
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niaEST:H3052B06-5"
/db_xref="taxon:10090"
/clone="H3052B06"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/note="Vector: pSPOR1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
BASE COUNT 157 a 168 c 181 g 122 t

ORIGIN
Alignment Scores:
Pred. No.: 3,34e-68 Length: 628
Score: 908.50 Matches: 186
Percent Similarity: 75.58% Conservative: 9
Best Local Similarity: 72.09% Mismatches: 12
Query Match: 14.04% Indels: 51
DB: 10 Gaps: 2
US-09-672-725c-25 (1-1281) x BG080311 (1-628)
QY 1024 SerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGluValVal 1043
Db 1 AGCACAGAGGCGTGAAGCCCTACTCTGTTAGAGGAAATGTAAATTTATGGAGTCCAG 60
QY 1044 PheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeuGluValLys 1063
Db 61 TTTAACTATCCACCCGACCCACATCCCACTGCTTCAGGGCGTGAAG 120

QY 1064 LysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrValValGln 1083
Db 121 AAGGCCAGAGCTTGGCCCTGGTGGCAGCAGTGCCTGTGGGAAGAGCACAGTGGTCCAG 180
QY 1084 LeuLeuGluArgPheTyrAppProLeuAlaGlySerValLeuLeuAspGlyLysGluIle 1103
Db 181 CTGCTCAGAGCGCTTCTACGACCCCGGATGATGATGATGATGATGATGATGATGATGAT 240
QY 1104 LysHisLeuAsnValGlnTTrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIle 1123
Db 241 AAGCACTGAATGTCAGTGGCTCCGAGCTCAGCTTGGCATTGTGCCAGGAGCCATT 300
QY 1124 LeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgValValSer 1143
Db 301 CTCCTTTCAGTGCAGCATTCGAGAGAACATCGCTATGGAGACACACCGCGCGTGTCT 360
QY 1144 HisGluGluLeuMetGlnAlaAlaLysGluAlaAsnIleHisPheIleGluThrLeu 1163
Db 361 CATGAGGAGATTGTGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
QY 1164 ProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGlnLys 1183
Db 421 CCTGAT----- 426
QY 1184 GlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeuAspGlu 1203
Db 426 ----- 426
QY 1204 AlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeuAspLysAla 1223
Db 427 -----GTTCTCCAGGAGCGCTGGACAAAGCC 453
QY 1224 ArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAlaAsp 1243
Db 454 AGGAGAGCGCCACCTGCATTGTCATCGCTCACCGCTGTCCACCATCCAGAACGCGGAC 513
QY 1244 LeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGlnGlnLeuLeu 1263
Db 514 TTGATCGTGTGTGATGAGAGCGGCAAGTCAAGGAGCAGCGCACCCACGAGCTGCTG 573
QY 1264 AlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLysArg 1281
Db 574 CGCAGAGGCGCATCTACTTCTCATG-----GTCCAGGCTGGAGCAAGCGC 621
RESULT 13 563 bp mRNA linear EST 23-OCT-2001
BI964198 ie66b04.y1 Melton Normalized Human Islet 4 M4-HIS 1 Homo sapiens
LOCUS cDNA 5' similar to SW:MDRI_HUMAN P08183 MULTIDRUG RESISTANCE
DEFINITION PROTEIN 1 ; mRNA sequence.
ACCESSION BI964198
VERSION BI964198.1 GI:16338603
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 563)
AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y., and Bowers, Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)

High quality sequence stop: 413.

Location/Qualifiers

FEATURES

source

1. .563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 136 a 132 c 153 g 142 t

ORIGIN

Alignment Scores:

Pred. No.: 1.39e-66 Length: 563
Percent Similarity: 889.00 Matches: 171
Best Local Similarity: 96.26% Conservative: 9
Query Match: 13.73% Mismatches: 7
DB: 10 Indels: 0 Gaps: 0

US-09-672-725c-25 (1-1281) x BI964198 (1-563)

QY 950 MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu 969
|||||
Db 3 ATGATGATTCTTCTATGCTGGATGTTCCGGTTTGGAGCTACTTGGTGGCACAATAA 62
QY 970 PheMetAsnPheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAla 989
|||||
Db 63 CTCATGAGCTTTGAGGATGTTCTGTAGTATTTTCAGCTGTGTGTTGGTCATGGCC 122
QY 990 ValGlyGlnValSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis 1009
|||||
Db 123 GTGGGGCAAGTCAGTTCATTGTCCTCGATGATGCCAAGCCAAATATCAGCAGCCAC 182
QY 1010 ValIleMetIleLeuGlySerProLeuIleAspSerTyrSerProHisGlyLeuLys 1029
:::|||||
Db 183 ATCATCATGATCATTTGAAACCCCTTGTATTGACAGCTACAGCAGGAGGAGGCTAATG 242
QY 1030 ProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArg 1049
|||||
Db 243 CCGAACACATTGGAAGAAATGTCACATTGTTGGTGAAGTGTATTCAACTATCCACCCGA 302
QY 1050 ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrLeuAla 1069
|||||
Db 303 CCGACATCCAGTCTTCAGGAGCTAGCTGAGGTGAAGAGGCCAGAGCGTGGCT 362
QY 1070 LeuValGlySerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyr 1089
|||||
Db 363 CTGGTGGCAGCAGTGGCTGTGGGAAGACGACACAGTGTGTCAGCTCCGAGCGGTTCTAC 422
QY 1090 AspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGln 1109
|||||

Db 423 GACCCCTGGCAGGGAAGTGCTGTGATGCGCAAGAAATAAAGCGACTGAATGTTTCAG 482
QY 1110 TrpLeuArgAlaHisIleuGlyIleValSerGlnGluProIleuPheAspCysSerIle 1129
|||||
Db 483 TGGCTCCGAGCACACCCGCGCATCGTGTCCAGGAGCCCATCTGTTTGACTGCAGCAT 542
QY 1130 AlaGluAsnIleAlaTyrGly 1136
|||||
Db 543 GCTGAGAACATTGCCTATGGA 563

RESULT 14

BI964218

LOCUS

DEFINITION

BI964218 563 bp mRNA linear EST 23-OCT-2001
cdna 5' similar to SW:MDRL_HUMAN P08183 MULTIDRUG RESISTANCE
PROTEIN 1, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 563)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)

TITLE

JOURNAL

COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)

High quality sequence stop: 367.

FEATURES

Source

1. .563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 136 a 132 c 153 g 142 t

ORIGIN

Alignment Scores:

Pred. No.: 1.39e-66 Length: 563

Score: 889.00 Matches: 171
Percent Similarity: 96.26% Conservative: 7
Best Local Similarity: 91.44% Mismatches: 9
Query Match: 13.73% Indels: 0
DB: 10 Gaps: 0

US-09-672-725C-25 (1-1281) x BI964218 (1-563)

QY 950 MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu 969
Db 3 ATGATGATATTTCCCTATGCTGGAGTGTTCGCGTTTGGAGCCTACTTGTGGCACATAAA 62
QY 970 PheMetAsnPheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAla 989
Db 63 CTCATGAGCTTTCAGCATGTTCTGTAGTATTTCAGCTGTGTCTTTGGTGGCATGGCC 122
QY 990 ValGlyGlnValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis 1009
Db 123 GTGGGCAAGTCACTATGCTCTGACTATGCCAAGGCCAAATATACAGCAGCCCCAC 182
QY 1010 ValIleMetIleLeuLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys 1029
Db 183 ATCATCATGATCATTTGAAAAACCCCTTTGATGTAGACGCTACAGCAGCGAAGCCCTAATG 242
QY 1030 ProAsnThrLeuGluGlyAsnValThrPheAsnGluValPheAsnTyrProThrArg 1049
Db 243 CCGAACACATTGGAAGGAATGTCATTTGGTGAAGTTGTATTCAACTATCCACCCGA 302
QY 1050 ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAla 1069
Db 303 CCGGACATCCCATGCTTTCAGGAGCTGAGCCTGGAGGTGAAGAAGGCCAGACGCTGGCT 362
QY 1070 LeuValGlySerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyr 1089
Db 363 CTGGTGGGCACAGTGGCTGTGGAGAGACACAGTGGTCCAGCTCTGGAGCGGTCTTAC 422
QY 1090 AspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGln 1109
Db 423 GACCCCTTGGCAGGAAGTGCCTGTGATGGCAAGAAATAAAGCGACTGATGATGTCAG 482
QY 1110 TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle 1129
Db 483 TGGCTCCGAGCACACCTGGGCATCGTGTCCAGGAGGCCCATCTGTTGACTGCACGATT 542
QY 1130 AlaGluAsnIleAlaTyrGly 1136
Db 543 GCTGAGAACATTGCCATGCA 563

RESULT 15

BM052878
LOCUS ie67e03.y3 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION CDNA 5' similar to SW:MDRL_HUMAN P08183 MULTIDRUG RESISTANCE
PROTEIN 1 ; mRNA sequence.

ACCESSION

BM052878

VERSION

BM052878.1 GI:16808672

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 563)

Leishka, I., Seearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,

Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,

Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas

, M., Gibbons, M., McCann, R., Cole, R., Tsagarelshvili, R., Williams, T.

, Jackson, Y., and Bowers, Y.

Endocrine Pancreas Consortium

Unpublished (2000)

CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
High quality sequence stop: 430.

FEATURES

source

1..563
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Not 1;
Site: 2; Sal 1; Starting library constructed using
SuperScript plasmid library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 136 a 132 c 153 g 142 t

ORIGIN

Alignment Scores:

Pred. No.: 1.39e-66 Length: 563
Score: 889.00 Matches: 171
Percent Similarity: 96.26% Conservative: 9
Best Local Similarity: 91.44% Mismatches: 7
Query Match: 13.73% Indels: 0
DB: 10 Gaps: 0

US-09-672-725C-25 (1-1281) x BM052878 (1-563)

QY 950 MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu 969
Db 3 ATGATGATATTTCCCTATGCTGGAGTGTTCGCGTTTGGAGCCTACTTGTGGCACATAAA 62

QY 970 PheMetAsnPheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAla 989
Db 63 CTCATGAGCTTTCAGCATGTTCTGTAGTATTTCAGCTGTGTCTTTGGTGGCATGGCC 122

QY 990 ValGlyGlnValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis 1009
Db 123 GTGGGCAAGTCACTATGCTCTGACTATGCCAAGGCCAAATATACAGCAGCCCCAC 182

QY 1010 ValIleMetIleLeuLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys 1029
Db 183 ATCATCATGATCATTTGAAAAACCCCTTTGATGTAGCAGCTACAGCAGGAAGCCCTAATG 242

QY 1030 ProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArg 1049
Db 243 CCGAACACATTGGAAGGAATGTCACATTTGGTGAAGTTGTATTCAACTATCCACCCGA 302

QY 1050 ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAla 1069
Db 303 CCGGACATCCCATGCTTTCAGGAGCTGAGCCTGGAGGTGAAGAAGGCCAGCAGCTGGCT 362

QY 1070 LeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyr 1089
Db 363 CTGGTGGGCACAGTGGCTGTGGGAGAGACACAGTGGTCCAGCTCTCTGGAGCGGTCTTAC 422

```

Qy 1090 AspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGln 1109
    |||||
Db 423 GACCCCTTGGCAGGGAAGTCTGCTTGGCAAGAAATAAACCGACTGAATGTTCAAG 482
    |||||
Qy 1110 TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle 1129
    |||||
Db 483 TGGCTCCGAGCACACCTGGGCATCGTGTCGCCAGGAGCCCATCTGTTTGACTGCAGCATT 542
    |||||
Qy 1130 AlaGluAsnIleAlaTyrGly 1136
    |||||
Db 543 GCTGAGAACATGCGCTATGGA 563
    |||||

```

Search completed: November 7, 2002, 07:29:20
 Job time : 2251.96 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 17:45:54 ; Search time 31.4943 Seconds
(without alignments)
4517.824 Million cell updates/sec

Title: US-09-672-725C-27

Perfect score: 6473

Sequence: 1 MDPEGRKGSAAENFWMGK.....LLAQGIYFMSVQAGAKR 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	6473	100.0	1281	22	AAE00310
2	6466	99.9	1281	22	AAE00309
3	6461	99.8	1281	22	AAE00308
4	6456	99.7	1281	22	AAE00303
5	6425.5	99.3	1280	22	AAE00304
6	6425.5	99.3	1280	22	AAE00304
7	5866	90.6	1280	18	AAW44073
8	5866	90.6	1280	21	AAW58186
9	5866	90.6	1280	22	AAW58186
10	5862	90.6	1280	22	AAW58186
11	5859.5	90.5	1283	22	AAW58186

12	5857	90.5	1280	21	AAW58187
13	5857	90.5	1280	22	AAW58186
14	5857	90.5	1280	22	AAE00306
15	5854	90.4	1280	8	AAE00306
16	5851	90.4	1280	11	AAE00306
17	5849.5	90.4	1279	22	AAE00307
18	5849.5	90.4	1279	22	AAE00307
19	5844	90.3	1280	15	AAE00307
20	5841	90.2	1280	14	AAE00307
21	5840.5	90.2	1280	14	AAE00307
22	5840.5	90.2	1279	19	AAW04347
23	5830	90.1	1280	19	AAW04347
24	5824	90.0	1280	19	AAW48997
25	5824	90.0	1280	19	AAW48997
26	5818	89.9	1280	19	AAW48999
27	5818	89.9	1280	19	AAW48999
28	5628	86.9	1272	21	AAW58189
29	5618	86.8	1272	21	AAW58189
30	5301	81.9	1276	14	AAW58189
31	5301	81.9	1276	14	AAW58189
32	5291.5	81.7	1275	21	AAW58188
33	5291.5	81.7	1275	21	AAW58188
34	2811.5	43.4	1275	20	AAW60409
35	2695	41.6	1313	22	AAW62594
36	2561	39.6	1279	22	ABW60234
37	2514	38.8	1349	20	ABW60234
38	2497	38.6	1302	22	ABW60234
39	2482	38.3	1334	20	ABW60234
40	2445	37.8	1408	19	AAW62871
41	2425	37.5	1286	21	AAW62871
42	2403.5	37.1	1333	22	AAW62871
43	2398	37.0	1243	21	AAW62871
44	2394	37.0	1320	22	ABW64922
45	2383	36.8	1254	21	AAW92173

ALIGNMENTS

RESULT 1

AAE00310

ID AAE00310 standard; Protein: 1281 AA.

AC AAE00310;

DT 13-JUN-2001 (first entry)

DE Dog P-glycoprotein (PGP) allelic variant (Genotype D).

Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;

MDRI; drug bioavailability; transgenic animal; genetic model.

OS Canis familiaris.

XX Key Location/Qualifiers

FT Misc-difference 25

FT /note= "Asn of GenotypeC substituted by Lys"

FT Misc-difference 197

FT /note= "His of GenotypeC substituted by Gln"

FT Misc-difference 329

FT /note= "Ser of GenotypeC substituted by Thr"

FT Misc-difference 1148

FT /note= "Met of GenotypeC substituted by Val"

FT WO200123540-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26767.

XX 28-SEP-1999; 99US-0156510.

XX (GENT-) GENTEST CORP.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX WO200123540-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26767.
XX 28-SEP-1999; 99US-0156510.
XX (GENT-) GENTEST CORP.
XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX WPI; 2001-235373/24.
XX N-PSDB; AAD03505.
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX for determining the bioavailability of drugs and for screening for dog
XX PGP inhibitors -
XX Claim 17; Page 99-102; 111pp; English.
XX The invention relates to dog P-glycoprotein (PGP) also referred
XX as multidrug transporter (MDR1) and nucleic acids encoding them.
XX The invention also includes fragments and biologically functional
XX variants of dog P-glycoprotein. PGP and their nucleic acids are
XX useful for determining the bioavailability of drugs and for
XX screening PGP inhibitors. They are useful for the diagnosis and
XX treatment of conditions characterised by PGP activity, by
XX reducing or increasing PGP activity in a cell. PGP nucleic acids
XX are used as oligonucleotide probes. Complements of PGP nucleic
XX acids are useful as antisense oligonucleotides, to induce a PGP
XX 'knockout' phenotype. They are used to prepare a non-human
XX transgenic animal, which are valuable as genetic models for
XX human diseases.
XX The present sequence is dog P-glycoprotein (PGP) allelic variant.
XX This sequence is also referred as Genotype B protein. The PGP
XX enzyme functions as an efflux pump exporting small molecules
XX across the cell membrane. This enzyme is a member of the ABC
XX transporter family.
XX Sequence 1281 AA;
Query Match 99.9%; Score 6466; DB 22; Length 1281;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1279; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPEGGRKGAENFWKMGKSKKKKKKKKKPTVSTFAMPRYSNWLDRMLVGTMAAII 60
DB 1 MDPEGGRKGAENFWKMGKSKKKKKKKKKPTVSTFAMPRYSNWLDRMLVGTMAAII 60
QY 61 HGAALPLMLVFGNMTDSPANAGISRNKTPPVIINESITNTTQHFINHLEEMTTVAYYY 120
DB 61 HGAALPLMLVFGNMTDSPANAGISRNKTPPVIINESITNTTQHFINHLEEMTTVAYYY 120
QY 121 SGIGAGVLVAAYIOVSPFWCLAAQRQILKIRKQFFHAIMRQIEIGWFDVHDVGNLRLTDD 180
DB 121 SGIGAGVLVAAYIOVSPFWCLAAQRQILKIRKQFFHAIMRQIEIGWFDVHDVGNLRLTDD 180
QY 181 VSKINEGIGDKIGMFFQSIATFTTGTIVGTRGWKTLVLTAISPVLGSAAIWAKILSS 240
DB 181 VSKINEGIGDKIGMFFQSIATFTTGTIVGTRGWKTLVLTAISPVLGSAAIWAKILSS 240
QY 241 FTDKELLAYAKAGAAVEVLAARTVIARGOKKELERYKNLEAKGIGIKKAITANIS 300
DB 241 FTDKELLAYAKAGAAVEVLAARTVIARGOKKELERYKNLEAKGIGIKKAITANIS 300
QY 301 IGAAFLLIYASALAFWYGTSLVLSSEYTGQVLTFFSVLIGAFSIGQASPSIEAFANA 360
DB 301 IGAAFLLIYASALAFWYGTSLVLSSEYTGQVLTFFSVLIGAFSIGQASPSIEAFANA 360
QY 361 RGAAYEIFKIIDNKPSIDSYSKSGHKPDNKGLEFNKVFHFSYPSRKEVKILKGLNLKVQ 420
DB 361 RGAAYEIFKIIDNKPSIDSYSKSGHKPDNKGLEFNKVFHFSYPSRKEVKILKGLNLKVQ 420

DB 361 RGAAYEIFKIIDNKPSIDSYSKSGHKPDNKGLEFNKVFHFSYPSRKEVKILKGLNLKVQ 420
QY 421 SGOTVALVGNSGCGKSTTVOLMQRLYDPTDGMVCIDGQDIRTINVRHLREITGVWSQEPV 480
DB 421 SGOTVALVGNSGCGKSTTVOLMQRLYDPTDGMVCIDGQDIRTINVRHLREITGVWSQEPV 480
QY 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKOR 540
DB 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKOR 540
QY 541 IAIARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
DB 541 IAIARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
QY 601 AGFDDGVIVVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKSSDALEMSPKDS 660
DB 601 AGFDDGVIVVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKSSDALEMSPKDS 660
QY 661 GSSLIKRRSTRRSIHAPQODRKLGTKEDLNENVPVSWFIRILKLNSTWMPYFVVGIFCA 720
DB 661 GSSLIKRRSTRRSIHAPQODRKLGTKEDLNENVPVSWFIRILKLNSTWMPYFVVGIFCA 720
QY 721 IINGGLQPAFSIIFSRIGIIFTRDEDPETKRONSNMFSVLFLVGLIISFITFFLQCFTEG 780
DB 721 IINGGLQPAFSIIFSRIGIIFTRDEDPETKRONSNMFSVLFLVGLIISFITFFLQCFTEG 780
QY 781 KAGEILTCLRMYFRSMRLRQDVSFDDPKNTTGALTTLRLANDAQAQVKAIGSLRAVITQ 840
DB 781 KAGEILTCLRMYFRSMRLRQDVSFDDPKNTTGALTTLRLANDAQAQVKAIGSLRAVITQ 840
QY 841 NTANLGTGIIISLIYQWQLTLLALLVPIIATAGVVEKMLSGOALKDKKELEGAGKTAT 900
DB 841 NTANLGTGIIISLIYQWQLTLLALLVPIIATAGVVEKMLSGOALKDKKELEGAGKTAT 900
QY 901 EAIENPRTVVSUTREQKFYMYAQSLOQVPRNSLRKAHIFGVFSFISITQAMMYFSYAGCFR 960
DB 901 EAIENPRTVVSUTREQKFYMYAQSLOQVPRNSLRKAHIFGVFSFISITQAMMYFSYAGCFR 960
QY 961 FGAYLVANEFMNFQDVLVFSIAIVFGAMAVGOVSSFADPYAKAKVSAAHVIMIEKSPLI 1020
DB 961 FGAYLVANEFMNFQDVLVFSIAIVFGAMAVGOVSSFADPYAKAKVSAAHVIMIEKSPLI 1020
QY 1021 DSYSPHGLKPNTLEGNTVNEVFNVPTRPDIPVLQGLSLEVKKGOTLALVSSCGCKST 1080
DB 1021 DSYSPHGLKPNTLEGNTVNEVFNVPTRPDIPVLQGLSLEVKKGOTLALVSSCGCKST 1080
QY 1081 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHLGIVSQEPILFDCSIAENIAYGDNRS 1140
DB 1081 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHLGIVSQEPILFDCSIAENIAYGDNRS 1140
QY 1141 VVSHEEIVQAAKEANIHFIETLPEKYNTRVGDKGTQLSGGOKQRIATARALVRPHILL 1200
DB 1141 VVSHEEIVQAAKEANIHFIETLPEKYNTRVGDKGTQLSGGOKQRIATARALVRPHILL 1200
QY 1201 LDEATSDALDTESEKVVQVALDKAREGRTTIVIAHRLSTIONADLIVFPQNGKVEHGTHQ 1260
DB 1201 LDEATSDALDTESEKVVQVALDKAREGRTTIVIAHRLSTIONADLIVFPQNGKVEHGTHQ 1260
QY 1261 QLLAQKGIYFMSVQAGAKR 1281
DB 1261 QLLAQKGIYFMSVQAGAKR 1281
RESULT 3
AAE00308
ID AAE00308 standard; Protein; 1281 AA.
XX
AC AAE00308;
XX
XX 13-JUN-2001 (first entry)
XX
DE Dog P-glycoprotein (PGP) allelic variant (Genotype A).
XX

KW Dog: P-glycoprotein allelic variant; PGP; multidrug transporter;
KW MDRL; drug bioavailability; transgenic animal; genetic model.
OS Canis familiaris.
FH Key Location/Qualifiers
FT Misc-difference 197
TT /note= "His of GenotypeC substituted by Gln"
XX WO200123540-A2.
PN 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26767.
XX 28-SEP-1999; 99US-0156510.
PR (GENT-) GENTEST CORP.
XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX WPI; 2001-235373/24.
DR N-PSDB; AAD03504.
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX Claim 17; Page 91-93; 11pp; English.
XX The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDRL) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein, PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
CC This sequence is also referred as Genotype A protein. The PGP
CC enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.
XX Sequence 1281 AA;
SQ Query Match 99.8%; Score 6461; DB 22; Length 1281;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1278; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MDPEGGRKGSAEKFWKMGKSKKKEKKKPTVSTFAMFRYSNWLDRMLVGTMAAII 60
Db 1 MDPEGGRKGSAEKFWKMGKSKKKEKKKPTVSTFAMFRYSNWLDRMLVGTMAAII 60
Qy 61 HGAALPLMLVFGNMTDSFANAGISRNKTPPVIINESITNTQHFINHLEEMTVYAYY 120
Db 61 HGAALPLMLVFGNMTDSFANAGISRNKTPPVIINESITNTQHFINHLEEMTVYAYY 120
Qy 121 SGIGAGVLVAAYIQVSWFLAAGRQILKIRKQFFHAIIMROEIGWFDVHGYELNRLTDD 180
Db 121 SGIGAGVLVAAYIQVSWFLAAGRQILKIRKQFFHAIIMROEIGWFDVHGYELNRLTDD 180
Qy 181 VSKINEGIDKIGMFFOSIAFTFTGFTVGTTRGKWLTLVLAISPVLGSLAAWAKILSS 240
Db 181 VSKINEGIDKIGMFFOSIAFTFTGFTVGTTRGKWLTLVLAISPVLGSLAAWAKILSS 240
Qy 241 FTDKELLAYAKAGAAEVLAARTVAFGGQKELBRYNKNLEEAIGIKKAITANIS 300
Db 241 FTDKELLAYAKAGAAEVLAARTVAFGGQKELBRYNKNLEEAIGIKKAITANIS 300

Db 241 FTDKELLAYAKAGAAEVLAARTVAFGGQKELBRYNKNLEEAIGIKKAITANIS 300
Qy 301 IGAAFLIIYASYALAFWYGTSLVLSSEYITIGQVLTFFSVLIGAFSIGQASPSIEAFANA 360
Db 301 IGAAFLIIYASYALAFWYGTSLVLSSEYISGOVLTFFSVLIGAFSIGQASPSIEAFANA 360
Qy 361 RGAAYEIFKIIDNKPSIDSYSGKHGKPDNKGLEFNKLVHFSYPSRKEVKILGLNKLKVQ 420
Db 361 RGAAYEIFKIIDNKPSIDSYSGKHGKPDNKGLEFNKLVHFSYPSRKEVKILGLNKLKVQ 420
Qy 421 SGQTVLVGNSGCGKSTTVQLMQRIDYPTDGMVCIQDQDRTINVRHLREITGVVSOEPV 480
Db 421 SGQTVLVGNSGCGKSTTVQLMQRIDYPTDGMVCIQDQDRTINVRHLREITGVVSOEPV 480
Qy 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKOR 540
Db 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKOR 540
Qy 541 IAIARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
Db 541 IAIARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
Qy 601 AGFDDGVIVEKGNHDELMEKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKDS 660
Db 601 AGFDDGVIVEKGNHDELMEKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKDS 660
Qy 661 GSSLIKRRSTRSIHAPQODRKLGTKEDLNENPVPSFWIRILKLNSTWEPYFVVGIFCA 720
Db 661 GSSLIKRRSTRSIHAPQODRKLGTKEDLNENPVPSFWIRILKLNSTWEPYFVVGIFCA 720
Qy 721 INGLOPAPSIISRIIGITRDEDPETKRONSMFSLFLVLGIISFIITFFLQGGTFFG 780
Db 721 INGLOPAPSIISRIIGITRDEDPETKRONSMFSLFLVLGIISFIITFFLQGGTFFG 780
Qy 781 KAGEILTRLRYMVRSMRLQDVSMFDDPKNTTGALTTRLANDAQAQVKAIGSLAVITQ 840
Db 781 KAGEILTRLRYMVRSMRLQDVSMFDDPKNTTGALTTRLANDAQAQVKAIGSLAVITQ 840
Qy 841 NIANGTGIISLIYGWQLTLLLAIVPIIAIAGVVENKMLSGQALKDKKELEGAGKIAT 900
Db 841 NIANGTGIISLIYGWQLTLLLAIVPIIAIAGVVENKMLSGQALKDKKELEGAGKIAT 900
Qy 901 EAIENFRVVSUTREQKPEYMAQSLQVYPYRNSLRKAHIFGVFSFISITQAMMYFSYAGCFR 960
Db 901 EAIENFRVVSUTREQKPEYMAQSLQVYPYRNSLRKAHIFGVFSFISITQAMMYFSYAGCFR 960
Qy 961 FGAYLVANEFMNFQDVLVLSAIVFGAMAVGVSSFADPYAKAKVSAAHVIMIEKSPLI 1020
Db 961 FGAYLVANEFMNFQDVLVLSAIVFGAMAVGVSSFADPYAKAKVSAAHVIMIEKSPLI 1020
Qy 1021 DSYSPHGLKPNTLEGNTVFNEVFNYPTRDIPVLQGLSLEVKKGOTLALVGSSCGCKST 1080
Db 1021 DSYSPHGLKPNTLEGNTVFNEVFNYPTRDIPVLQGLSLEVKKGOTLALVGSSCGCKST 1080
Qy 1081 VVOLLERFDPLAGSVLIDGKEIKHLNVQWLAHILGIVSQEPILFDCSIAENIAYGDNRS 1140
Db 1081 VVOLLERFDPLAGSVLIDGKEIKHLNVQWLAHILGIVSQEPILFDCSIAENIAYGDNRS 1140
Qy 1141 VVSHEEIVQAQAEANIHHPIETLPEKYNTRVGDGKTQLSGGOKORIAARALVRPHILL 1200
Db 1141 VVSHEEIVQAQAEANIHHPIETLPEKYNTRVGDGKTQLSGGOKORIAARALVRPHILL 1200
Qy 1201 LDEATSDALDTESEKVVQEAQALDKARSGRTCTIVIAHRLSTIQNADLIIVFQNGVKEHGTHQ 1260
Db 1201 LDEATSDALDTESEKVVQEAQALDKARSGRTCTIVIAHRLSTIQNADLIIVFQNGVKEHGTHQ 1260
Qy 1261 QLLAOKGIYFSMVSVQAGAKR 1281
Db 1261 QLLAOKGIYFSMVSVQAGAKR 1281
RESULT 4
AAE00303

RESULT 5

AB81068
ID AAB01068 standard; Protein; 1280 AA.XX
AC AAB81068;XX
DT 25-JUN-2001 (first entry)XX
DE Dog P-glycoprotein SEQ ID 7.XX
KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW efflux pump; dog.XX
OS Canis familiaris.XX
PN WO200123565-A1.XX
PD 05-APR-2001.XX
PF 28-SEP-2000; 2000WO-US26592.XX
PR 28-SEP-1999; 99US-0156921.XX
PR 12-OCT-1999; 99US-0158818.XX
PA (GENT-) GENTEST CORP.XX
PI Stocker PJ, Steimel-Crespi DT, Crespi CL;XX
DR WPI; 2001-316136/33.XX
PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell.XX
PS Claim 8; Page 73-76; 84pp; English.XX
CC This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDR1 is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents
CC P-glycoprotein from a dog.XX
SQ Sequence 1280 AA;Query Match 99.3%; Score 6425.5; DB 22; Length 1280;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1273; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 1 MDPEGRKGSAAENFWKMGKSKKKKKKPTVSTFAMFRYSNWLDRMLVGTMAAII 60

DB 1 MDPEGRKGSAAENFWKMGKSKKKKKKPTVSTFAMFRYSNWLDRMLVGTMAAII 59

QY 61 HGAALPLMLVFGNWTDSFANAGISRNTFPVVIINESITNTQHFNLHLEEMTYAYYY 120

DB 60 HGAALPLMLVFGNWTDSFANAGISRNTFPVVIINESITNTQHFNLHLEEMTYAYYY 119

QY 121 SGIGAGVLVAAYIQVSEFWCLAAQRQILKIRKQFFHAIMRQEGWFDVHDVDELNRLTDD 180

DB 120 SGIGAGVLVAAYIQVSEFWCLAAQRQILKIRKQFFHAIMRQEGWFDVHDVDELNRLTDD 179

QY 181 VSKINEGIDKGMFFQSIATFTFTGIVGTGRGWKLTVLVAISPVGLSAAIWKILSS 240

DB 180 VSKINEGIDKGMFFQSIATFTFTGIVGTGRGWKLTVLVAISPVGLSAAIWKILSS 239

QY 241 FTDKELLAYAKAGAAVEVLAAIIRTVIATGGQKKELERNKNLEAAKGIGIKKAITANIS 300
DB 240 FTDKELLAYAKAGAAVEVLAAIIRTVIATGGQKKELERNKNLEAAKGIGIKKAITANIS 299
QY 301 IGAFLLIYASYALAFWYGTSLVLSSEYITIGQVLTFFVSVLIGAFSIGQASPSIEAFANA 360
DB 300 IGAFLLIYASYALAFWYGTSLVLSSEYITIGQVLTFFVSVLIGAFSIGQASPSIEAFANA 359
QY 361 RGAAYEIFKIIDNKPSIDSYSKSGHKPDNIKGNLEFNKNHFSYPSRKEVKILGMLKVQ 420
DB 360 RGAAYEIFKIIDNKPSIDSYSKSGHKPDNIKGNLEFNKNHFSYPSRKEVKILGMLKVQ 419
QY 421 SGQTVALVNSGGCKSTTVOLMQRLYDPTDGMVCIDQDIRTINVRLHREITGVVSOEYV 480
DB 420 SGQTVALVNSGGCKSTTVOLMQRLYDPTDGMVCIDQDIRTINVRLHREITGVVSOEYV 479
QY 481 LFATTIAENIRYGRNVTMDIEIEKAVKEANAYDFIMKLPNKFTLVGERGAQLSGGQKOR 540
DB 480 LFATTIAENIRYGRNVTMDIEIEKAVKEANAYDFIMKLPNKFTLVGERGAQLSGGQKOR 539
QY 541 IATARALVRNP KILLDEAT SALDTESEAVVQVALDKARKGRITTVIAHRLSTVRNADVI 600
DB 540 IATARALVRNP KILLDEAT SALDTESEAVVQVALDKARKGRITTVIAHRLSTVRNADVI 599
QY 601 AGFDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNETELEENATGESKESDALEMSPKDS 660
DB 600 AGFDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNETELEENATGESKESDALEMSPKDS 659
QY 661 GSSLIKKRRSTRRSIHAPQGGDRKLGTKEDLNENVPYSFWRILKLNSTWEPYFVVGIFCA 720
DB 660 GSSLIKKRRSTRRSIHAPQGGDRKLGTKEDLNENVPYSFWRILKLNSTWEPYFVVGIFCA 719
QY 721 IINGLOPAFSIIPRSIIGTITREDDETPETKRONSMFVFLVGLIISFTIFFLQGTFFG 780
DB 720 IINGLOPAFSIIPRSIIGTITREDDETPETKRONSMFVFLVGLIISFTIFFLQGTFFG 779
QY 781 KAGELTKRLRYMFRSMRLQDVSNFDDPKNTTGALTTRLANDAAQVKAIGSLRAVITQ 840
DB 780 KAGELTKRLRYMFRSMRLQDVSNFDDPKNTTGALTTRLANDAAQVKAIGSLRAVITQ 839
QY 841 NIANLGTGIIISLIYQWLTLLLLAIVIPIIAIAAGVEMKMLSGQALKDKKELEGAGKIAT 900
DB 840 NIANLGTGIIISLIYQWLTLLLLAIVIPIIAIAAGVEMKMLSGQALKDKKELEGAGKIAT 899
QY 901 EAIENFTVSLTREQFEYMAQSLQVPPYNSLRKAHIFGVSPSITQAMMIFSYAGCFR 960
DB 900 EAIENFTVSLTREQFEYMAQSLQVPPYNSLRKAHIFGVSPSITQAMMIFSYAGCFR 959
QY 961 FGAYLVANEFNFQDVLIVFSAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPLI 1020
DB 960 FGAYLVANEFNFQDVLIVFSAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPLI 1019
QY 1021 DSYSPHGLKPNLTGCVTFNEVFNYPTRPDIPVQLGLEVKKGQTLALVGGSGCKST 1080
DB 1020 DSYSPHGLKPNLTGCVTFNEVFNYPTRPDIPVQLGLEVKKGQTLALVGGSGCKST 1079
QY 1081 VVQLLERYDPLAGSVLDIGKEIKHLNVQWLRHLGVSQEPILFDCSIAENIAYGNSR 1140
DB 1080 VVQLLERYDPLAGSVLDIGKEIKHLNVQWLRHLGVSQEPILFDCSIAENIAYGNSR 1139
QY 1141 VVSHEEIVQAAKEANIHFIETLPEKYNTRYGDKGTQLSGGQKORIAIARALVRQPHILL 1200
DB 1140 VVSHEEIVQAAKEANIHFIETLPEKYNTRYGDKGTQLSGGQKORIAIARALVRQPHILL 1199
QY 1201 IDEATSALDTESEKVVQEQALDKAREGRTCVIAHRLSTIQNADLIIVFQNGKVEHGTQ 1260
DB 1200 IDEATSALDTESEKVVQEQALDKAREGRTCVIAHRLSTIQNADLIIVFQNGKVEHGTQ 1259
QY 1261 QLLAOKGIYFESMVSVQAGAKR 1281
DB 1260 QLLAOKGIYFESMVSVQAGAKR 1280

RESULT 6
AAE00304
ID AAE00304 standard; Protein; 1280 AA.
XX AC AAE00304;
XX DT 13-JUN-2001 (first entry)
XX DE Dog P-glycoprotein (PGP) #2.
XX KW Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
XX KW drug bioavailability; transgenic animal; genetic model.
XX OS Canis familiaris.
XX PN WO200123540-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26767.
XX PR 28-SEP-1999; 99US-0156510.
XX PA (GENT-) GENTEST CORP.
XX PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX DR WPI; 2001-235373/24.
XX DR N-PSDB; AAD03489.
XX PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX PT for determining the bioavailability of drugs and for screening for dog
XX PT PGP inhibitors -
XX PS Example 2; Page 72-75; lllpp; English.
XX CC The invention relates to dog P-glycoprotein (PGP) also referred
XX CC as multidrug transporter (MDR1) and nucleic acids encoding them.
XX CC The invention also includes fragments and biologically functional
XX CC variants of dog P-glycoprotein. PGP and their nucleic acids are
XX CC useful for determining the bioavailability of drugs and for
XX CC screening PGP inhibitors. They are useful for the diagnosis and
XX CC treatment of conditions characterised by PGP activity, by
XX CC reducing or increasing PGP activity in a cell. PGP nucleic acids
XX CC are used as oligonucleotide probes. Complements of PGP nucleic
XX CC acids are useful as antisense oligonucleotides, to induce a PGP
XX CC 'knockout' phenotype. They are used to prepare a non-human
XX CC transgenic animal, which are valuable as genetic models for
XX CC human diseases.
XX CC The present sequence is dog P-glycoprotein (PGP). The
XX CC PGP enzyme functions as an efflux pump exporting small molecules
XX CC across the cell membrane. This enzyme is a member of the ABC
XX CC transporter family.
SQ Sequence 1280 AA;
Query Match 99.3%; Score 6425.5; DB 22; Length 1280;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1273; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
QY 1 MDPGGKGAENKFNKMGKKKKKKKKKPTVSTVFAMFRYSNWLDRMLVGTWMAII 60
DB 1 MDPGGKGAENKFNKMGKKK-KKEKKKKPTVSTVFAMFRYSNWLDRMLVGTWMAII 59
QY 61 HGAALPLMLVFCNWTDSFANAGISRNKTFPVIINESITNNTQHFTHLHEEMTYAYYY 120
DB 60 HGAALPLMLVFCNWTDSFANAGISRNKTFPVIINESITNNTQHFTHLHEEMTYAYYY 119
QY 121 SGIGAGLVAAAYIQVSFWCLAGRQLTKRKQFFHAIMRQETGWEDVDVHGLNRLTDD 180
DB 120 SGIGAGLVAAAYIQVSFWCLAGRQLTKRKQFFHAIMRQETGWEDVDVHGLNRLTDD 179
QY 181 VSKINEGIDKGMFFQSIATFTTGFIVGTGPKWKLTVILAISPVLGSLAAIWAIISS 240

DB 180 VSKINEGIDKGMFFQSIATFTTGFIVGTGPKWKLTVILAISPVLGSLAAIWAIISS 239
QY 241 FTDKELLAYAKAGAAVEVLAAITVIAFGQKKELERNKNLEAKGIGIKKAITANIS 300
DB 240 FTDKELLAYAKAGAAVEVLAAITVIAFGQKKELERNKNLEAKGIGIKKAITANIS 299
QY 301 IGAFLLIYASALAFWYGTSLVLSSEYITIGQVLTFFVFFSVLIGAFSIGQASPIEAFANA 360
DB 300 IGAFLLIYASALAFWYGTSLVLSSEYITIGQVLTFFVFFSVLIGAFSIGQASPIEAFANA 359
QY 361 RGAAYEIFKIIDNKPSIDSYSGKHDPDNIKGNLEPKNVHFSYPSRKEVKILKGLNKVQ 420
DB 360 RGAAYEIFKIIDNKPSIDSYSGKHDPDNIKGNLEPKNVHFSYPSRKEVKILKGLNKVQ 419
QY 421 SGOVALVGNCGCKSTTVQLMORLYDPTDGMVCIDGQDIRTINVHRLREITGVVSOEPV 480
DB 420 SGOVALVGNCGCKSTTVQLMORLYDPTDGMVCIDGQDIRTINVHRLREITGVVSOEPV 479
QY 481 LFATTIAENIRYGRNVMTDEIEKAYKEANAYDFIMKLPNKFDTLVGERGAOLSGQKQR 540
DB 480 LFATTIAENIRYGRNVMTDEIEKAYKEANAYDFIMKLPNKFDTLVGERGAOLSGQKQR 539
QY 541 IATARALVRNPKILLDEATSAIDTSEAVQVVALDKARKGRTTIVIAHRLSTVRNADY 600
DB 540 IATARALVRNPKILLDEATSAIDTSEAVQVVALDKARKGRTTIVIAHRLSTVRNADY 599
QY 601 AGFDDGVIIVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKDS 660
DB 600 AGFDDGVIIVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKDS 659
QY 661 GSSLIKRRSTRRSIHAPQGDRLKGTEDLNENPVPSFWRLKLNSTENWPFYVVGIFCA 720
DB 660 GSSLIKRRSTRRSIHAPQGDRLKGTEDLNENPVPSFWRLKLNSTENWPFYVVGIFCA 719
QY 721 IINGGLQPAFSITFSRIIGIFTRDEDPETKRONSNMFSVLFLVLGIISFIFFLGQFTFG 780
DB 720 IINGGLQPAFSITFSRIIGIFTRDEDPETKRONSNMFSVLFLVLGIISFIFFLGQFTFG 779
QY 781 KAGEILTKRLRYMFRSLRQDVSWFDDPKNTGALTTRLANDAAQVKGAGISRLAVITQ 840
DB 780 KAGEILTKRLRYMFRSLRQDVSWFDDPKNTGALTTRLANDAAQVKGAGISRLAVITQ 839
QY 841 NTANLGTGIIISLIYGVQTLTLLLAIVPIIATAGVVMKMLSGQALKDKKELEGAGKIAT 900
DB 840 NTANLGTGIIISLIYGVQTLTLLLAIVPIIATAGVVMKMLSGQALKDKKELEGAGKIAT 899
QY 901 EAIENFRVTSUTREQKFEYMAQSIQVYPYNSLRKAHIFGVSPSITQAMMYFSYAGCFR 960
DB 900 EAIENFRVTSUTREQKFEYMAQSIQVYPYNSLRKAHIFGVSPSITQAMMYFSYAGCFR 959
QY 961 FGAYLVANEFMNFQDVLVLSAIVFEGAMAVGVSSFADYAKAKVSAARHVIIEKSPLI 1020
DB 960 FGAYLVANEFMNFQDVLVLSAIVFEGAMAVGVSSFADYAKAKVSAARHVIIEKSPLI 1019
QY 1021 DSYSPHGLKPNTEGNTVFNEVFNYPTRDIPVLOGLSLEVKGQTLALVSSGCGKST 1080
DB 1020 DSYSPHGLKPNTEGNTVFNEVFNYPTRDIPVLOGLSLEVKGQTLALVSSGCGKST 1079
QY 1081 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSIAENIAYGNSR 1140
DB 1080 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSIAENIAYGNSR 1139
QY 1141 VVSHEEIVQAAKEANIHHFIETLPEKYNTRYGDKGTQLSGGQKORIAIARALVRQPHLL 1200
DB 1140 VVSHEEIVQAAKEANIHHFIETLPEKYNTRYGDKGTQLSGGQKORIAIARALVRQPHLL 1199
QY 1201 LDEATSAIDTSEKVVQVQALDKAREGRTICIVIAHRLSTIQNADLIVVFQNGKVEHGTQ 1260
DB 1200 LDEATSAIDTSEKVVQVQALDKAREGRTICIVIAHRLSTIQNADLIVVFQNGKVEHGTQ 1259
QY 1261 OLLAQKGIYFSMVSVOAGAKR 1281
|||||:|||||

Db 1260 QLLAQKGIYFSMISVQAGAKR 1280

RESULT 7

AAW44073
ID AAW44073 standard; protein; 1280 AA.

XX AAW44073;

XX 26-JUN-1998 (first entry)

XX Human multidrug resistance P-glycoprotein MDRI.

XX Human; multidrug resistance P-glycoprotein; MDRI; prokaryotic homologue.

XX Homo sapiens.

XX Key Location/Qualifiers
XX Region 1..640

XX FT /note= "MDRI-N from Fig 1"
XX Region 641..1280

XX FT /note= "MDRI-C from Fig 1"

XX W09740160-A1.

XX PN 30-OCT-1997.

XX PF 24-APR-1997; 97WO-N000216.

XX PR 24-APR-1996; 96EP-0201094.

XX PA (UYGR-) RIJKSUNIV GRONINGEN.

XX PI Bolhuis H, Konings WN, Van Veen HW, Venema K;

XX WPI; 1997-535844/49.

XX Prokaryotic homologue of human multiple drug resistance protein -
XX used to screen for compounds that inhibit, or avoid, drug resistance
XX Claim 10; Fig 1; 35pp; English.

XX The present invention describes a recombinant or isolated nucleic acid
XX (1), derived from a prokaryotic gene, which encodes at least a specific
XX and/or functional part of a transporter protein (TP), or its
XX derivatives, which has functional and/or structural similarity with the
XX P-glycoprotein (PG) encoded by the human multidrug resistance
XX (MDR1) gene. The present sequence represents the human MDR1 protein,
XX derived from MDR1-N and MDR1-C as shown in the specification in
XX figure 1. (1) is used to express recombinant proteins; its fragments
XX are also useful as probes and primers for detection and amplification
XX of related DNA. The protein produced, or cells expressing them, are
XX used to determine if substances can inhibit, or avoid, MDR proteins,
XX and in a screening method for identifying compounds that inhibit
XX transport of cytotoxic substances from cells. Also, cells with a
XX transmembrane protein, especially where expressed from (1), can provide
XX (additional) MDR, particularly for use as a model system to study
XX mechanisms of action of PG.

XX Sequence 1280 AA;

Query Match 90.6%; Score 5866; DB 18; Length 1280;
Best Local Similarity 90.8%; Pred. No. 0;
Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;

Qy 1 MDPEGGKGS-A-EKNFWKMKKKKKKKKPTVSTFAMFRYSNWLDRMLVGTMAAI 59

Db 1 MDLEGRNGGAKKFNFLNKS-EKDKKKKPTVSVFMSFRYSNWLKLYMVGTAAI 59

Qy 60 IHGAALPLMLLVFGNMTDSFANAGISRNKTFPVINESITNTQHFHINLEEMTYAY 119

Db 60 IHGAGLPLMLLVFGEMTDIFANAG-NLEDLMSNITNSDNDTGTFMN-LEEDMTRYAY 117

Qy 120 YSGIGAGVLVAAYIQVSWFCLAAGRQILKIRKQFFHAIMRQEIQWDFVDHVGELNRLTD 179
Db 118 YSGIGAGVLVAAYIQVSWFCLAAGRQIHKIRKQFFHAIMRQEIQWDFVDHVGELNRLTD 177
Qy 180 DVSKINEGIGDKIGMFFQSTATPTFTGIVGTRGWKLTVLILAIISPVLGLSAAIWAKILS 239
Db 178 DVSKINEGIGDKIGMFFQSMATFTFTGIVGTRGWKLTVLILAIISPVLGLSAAIWAKILS 237
Qy 240 SFTDKELLAYAKAGAVAEVLAARTVIAFGGOKKELERYKNLEAKRIGIKKAITANI 299
Db 238 SFTDKELLAYAKAGAVAEVLAARTVIAFGGOKKELERYKNLEAKRIGIKKAITANI 297
Qy 300 SIGAAFLIIVASYALAFWYGTSLVLSSEYTIQVLTFFSVLIGAFSIGOASPSIEAFAN 359
Db 298 SIGAAFLIIVASYALAFWYGTSLVLSSEYTIQVLTFFSVLIGAFSIGOASPSIEAFAN 357
Qy 360 ARGAAEYEFKIIDNKPSIDSYSGHKPDNIKGNLEPKNFHVSYPSPRKEVKILKGLMLKV 419
Db 358 ARGAAEYEFKIIDNKPSIDSYSGHKPDNIKGNLEPKNFHVSYPSPRKEVKILKGLMLKV 417
Qy 420 QSGOTVALVNGSGCGKSTTVOLMORLYDPTDGMVCIDGQDITRTINVRHLREITGVVSOEP 479
Db 418 QSGOTVALVNGSGCGKSTTVOLMORLYDPTDGMVCIDGQDITRTINVRHLREITGVVSOEP 477
Qy 480 VLFATTIAENIRYGRNVTMDIEIKAVKEANAYDFIMKLPNKFDTLVGERGAOLSGGOKQ 539
Db 478 VLFATTIAENIRYGRNVTMDIEIKAVKEANAYDFIMKLPNKFDTLVGERGAOLSGGOKQ 537
Qy 540 RIATARALVRNPKILLDEATSAIDTESEAVVQVALDKARKGRITTVIAHRLSVRNADV 599
Db 538 RIATARALVRNPKILLDEATSAIDTESEAVVQVALDKARKGRITTVIAHRLSVRNADV 597
Qy 600 IAGDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELEENATGESKESDALEMSPKD 659
Db 598 IAGDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELEENATGESKESDALEMSND 657
Qy 660 SGSSLIKRSTRSRSHAPQGGODKLTGKTENENPVPSFWRILKLNSTENPYPVWVGIFC 719
Db 658 SRSSLIKRSTRSRVSGAQDRKLTKEALDESIPPVFSWRIMKLNLTENPYPVWVGIFC 717
Qy 720 AINGLQPAFISFIISRIIGITFDEDEPETKRONSMFSLVFLVIGIISFTFFLQGFTE 779
Db 718 AINGLQPAFISFIISRIIGITFDEDEPETKRONSMFSLVFLVIGIISFTFFLQGFTE 777
Qy 780 KRAGELKRLRYMVFRLMQLQDVSFDDPKNTGALTTRLANDAAOVKGAIGSLAVIT 839
Db 778 KRAGELKRLRYMVFRLMQLQDVSFDDPKNTGALTTRLANDAAOVKGAIGSLAVIT 837
Qy 840 QNIANLGTGIISLIYQWLTLLILAIPIIATAGVVMKMLSGOALKDKKKELEGAGKIA 899
Db 838 QNIANLGTGIISLIYQWLTLLILAIPIIATAGVVMKMLSGOALKDKKKELEGAGKIA 897
Qy 900 TEAIENFRTVSLTREQKFEYMAOSLOVPRNSLRKAHIFGVFSFTQAMMYSYAGCF 959
Db 898 TEAIENFRTVSLTREQKFEYMAOSLOVPRNSLRKAHIFGVFSFTQAMMYSYAGCF 957
Qy 960 RFGAYLVANEFNFODVLLVFSIAIVFGAMAVGOVSSPADYAKAKVSAHVIMIEKSPL 1019
Db 958 RFGAYLVANEFNFODVLLVFSIAIVFGAMAVGOVSSPADYAKAKVSAHVIMIEKSPL 1017
Qy 1020 IDSYSPHGLKPNLTLEGNTVFNEVFNTPRPDIPVLOGLSLEVKKGTALVVGSGGCKS 1079
Db 1018 IDSYSPHGLKPNLTLEGNTVFNEVFNTPRPDIPVLOGLSLEVKKGTALVVGSGGCKS 1077
Qy 1080 TVVQLLERYDPLAGSVLIDGKEIKHLNVQWLAHGLIVSQEPILFDCSIAENTAYGNS 1139
Db 1078 TVVQLLERYDPLAGSVLIDGKEIKHLNVQWLAHGLIVSQEPILFDCSIAENTAYGNS 1137
Qy 1140 RVVSEEEIVQAAKEANIHFIETLPEKNTRVGDKGTOLSGGOKORIAIARALVRQPHIL 1199
Db 1138 RVVSEEEIVQAAKEANIHFIETLPEKNTRVGDKGTOLSGGOKORIAIARALVRQPHIL 1197
Qy 1200 LLDDEATSAIDTESEKVVQVALDKAREGRTCTVIAHRLSTIQNADLIIVFQNGKVEHGT 1259

Db 1198 LLEDEATSDTSEKVVQEQALDKAREGRTCIIVIAHRLSTIQNADLLVVFQNGRVEKHGTH 1257
QY 1260 QQLLAQKGIYFSWVSVQAGAKR 1281
Db 1258 QQLLAQKGIYFSWVSVQAGTKR 1279

RESULT 8
AA58186
ID AA58186 standard; Protein; 1280 AA.
AC AA58186;
XX
DT 14-MAR-2000 (first entry)
DE Human wild-type multidrug resistance-1 (MDR-1) protein.
XX
KW Multidrug resistance; MDR-1; P-glycoprotein;
KW transmembrane efflux pump; haematopoietic stem cell; transduction;
KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW gene therapy; gene replacement; genetic defect; thalassaemia;
KW Gaucher's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW cytokine; wild-type.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 185
FT /note= "this residue is Val in a mutant MDR-1
FT (AA58187)"
XX
PN WO9961589-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11825.
XX
PR 28-MAY-1998; 98US-0086988.
XX
PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Sorrentino B, Bunting K;
XX
DR WPI; 2000-072615/06.
DR N-PSDB; AAZ49332.
XX
PT Ex vivo expansion of hematopoietic stem cells transduced with a
PT sequence encoding human multidrug resistance-1, used for bone marrow
PT transplantation -
XX
PS Claim 10; Page 71-79; 113pp; English.
XX
CC This sequence represents human wild-type multidrug
CC resistance protein MDR-1. MDR-1 is a transmembrane efflux
CC pump, responsible for the export of drugs from certain
CC cells, particularly cancer cells. Wild-type MDR-1 shows
CC increased resistance to etoposide and decreased resistance to vinca
CC alkaloids compared with a mutant form (AA58187) where the Gly at
CC position 185 is replaced by Val. The invention relates to transducing
CC haematopoietic stem cells with nucleic acid encoding an MDR protein
CC and culturing the modified cells. The modified haematopoietic stem
CC cells are useful in bone marrow transplantation (to reconstitute
CC haematopoietic systems in patients who have undergone chemotherapy or
CC radiation therapy) and in ex vivo gene therapy of genetic defects in
CC cells derived from haematopoietic stem cells, e.g., thalassaemia,
CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
CC cells can also be used to identify factors involved in regulating
CC proliferation and differentiation in haematopoietic stem cells.
CC Haematopoietic stem cells that express MDR-1 will be protected against
CC chemotherapeutic agents, so can be engrafted while the patient is
CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC provides sufficient cells to permit standard biochemical analysis.

CC Overexpression of MDR-1 allows cytokine-driven expansion of
CC haematopoietic stem cells by at least 10-fold compared with a maximum
CC of 4-fold in known procedures.
XX
SQ Sequence 1280 AA;
Query Match 90.6%; Score 5866; DB 21; Length 1280;
Best Local Similarity 90.8%; Pred. No. 0;
Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;
QY 1 MDPGGKRGSA-EKNFWKMGKKKKKKKPTVSTFAMFRYSNWLDRYMLVGMMAAI 59
Db 1 MDLEGRNGGAKKKKFKLNKS-EXDKKKKPTVSFMSFRYSNWLDRYMLVGMVGLAAI 59
QY 60 IHGAALPLMMLVFGNMTDSPANAGISRNTFPVVIINESITNTNQHFINHLEEMTYAYY 119
Db 60 IHGAGLPLMMLVFGEMTDIFANAG-NLEDLSNITNRSINDTGTFFMN-LEEDMTYAYY 117
QY 120 YSGIGAGVLVAAYIQVSFWCLAAAGROILLKIRKOFFHAIMRQEIQWDFVHDVGLNRLTD 179
Db 118 YSGIGAGVLVAAYIQVSFWCLAAAGROILLKIRKOFFHAIMRQEIQWDFVHDVGLNRLTD 177
QY 180 DVSKINEGIGDKIMPFQSIATFTFTGFTRGWKLTLVLAISVPLGISAIAWAKILS 239
Db 178 DVSKINEGIGDKIMPFQSMATFTFTGFTRGWKLTLVLAISVPLGISAIAWAKILS 237
QY 240 SFTDKELLAYAKAGAVAEVLAARTVIAFGGQKKELERYNKNLEAKGIGIKKAITANI 299
Db 238 SFTDKELLAYAKAGAVAEVLAARTVIAFGGQKKELERYNKNLEAKGIGIKKAITANI 297
QY 300 SIGAAFLLIYASALAFWYGTSLVLSSEYTIQVLTVFVSVLIGAFSIOQASPSIEAFAN 359
Db 298 SIGAAFLLIYASALAFWYGTSLVLSSEYTIQVLTVFVSVLIGAFSIOQASPSIEAFAN 357
QY 360 ARGAAYEIIFKIIDNKPSIDSYSKSGHKPDNIKGNLFKNVHFSYPSRKEVKILKGLNLKV 419
Db 358 ARGAAYEIIFKIIDNKPSIDSYSKSGHKPDNIKGNLFKNVHFSYPSRKEVKILKGLNLKV 417
QY 420 QSGQTVALVGNCGCKSTTVQLMQRLYDPTDGMVCIDGQDIRTINVVRHREITGVVSOBP 479
Db 418 QSGQTVALVGNCGCKSTTVQLMQRLYDPTDGMVCIDGQDIRTINVVRHREITGVVSOBP 477
QY 480 VLFATTIAENIRYGRNVMTDETEKAVKEANAYDFTMKLPNKFDTLVGSRGAQLSGGOK 539
Db 478 VLFATTIAENIRYGRNVMTDETEKAVKEANAYDFTMKLPNKFDTLVGSRGAQLSGGOK 537
QY 540 RTAARALVRNPKILLDEATSAIDTSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADV 599
Db 538 RTAARALVRNPKILLDEATSAIDTSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADV 597
QY 600 IAGFDDGVIVEKGNHDELMEKGIYFKLVMTQTRGNETELENAATESKESDALEMSPKD 659
Db 598 IAGFDDGVIVEKGNHDELMEKGIYFKLVMTQTRGNETELENAATESKESDALEMSPKD 657
QY 660 SGSSLLIKRSTRRSIHAPOGQDKLTGKEDLNENPVPSFWRLKLNSTEWPFYVVGIFC 719
Db 658 SRSSLLIKRSTRRSVRSQAQRKLTKEALDESIPPVSFWRLKLNSTEWPFYVVGIFC 717
QY 720 AIINGLOPAFAIIFSKIIGVTRIDDPETKQNSNMFSLVFLVGLIISFIFTLQGTFF 779
Db 718 AIINGLOPAFAIIFSKIIGVTRIDDPETKQNSNMFSLVFLVGLIISFIFTLQGTFF 777
QY 780 GKAGEILLKRLRYMFRSMRLRODVSWFDDPKNTTGALTTRLANDAAQVKAIGSLAVIT 839
Db 778 GKAGEILLKRLRYMFRSMRLRODVSWFDDPKNTTGALTTRLANDAAQVKAIGSLAVIT 837
QY 840 QNIANLGTGIISLYGWLTLLLAIIVPIIAIGVEMKMLSGOALKDKKLEAGAKIA 899
Db 838 QNIANLGTGIISLYGWLTLLLAIIVPIIAIGVEMKMLSGOALKDKKLEAGAKIA 897
QY 900 TEAIENFRVTWLTREQFEYMYAQSLQVPVYNSLRKAHIFGVFSFISITQAMMYSYACGF 959
Db 898 TEAIENFRVTWLTREQFEYMYAQSLQVPVYNSLRKAHIFGVFSFISITQAMMYSYACGF 957

QY	960	REFAYLVANEFNFQDVLVLSAIVFAGAMAVGOVSSFPADYAKAVSAAHVIMITEKSPL	1019
DB	958	RFGAYLVVAHLKMSFEDVLVLSAIVFAGAMAVGOVSSFPADYAKAKISAHHIIMIEKTP	1017
QY	1020	IDSYSRPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS	1079
DB	1018	IDSYSTLEGLMPNTLEGNVTFGEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS	1077
QY	1080	TVVOLLERYDPLAGSVLIDGKEIKHLNVQWLRHAHLGIVSQEPILPDCSIAENIAYGNS	1139
DB	1078	TVVOLLERYDPLAGSVLIDGKEIKHLNVQWLRHAHLGIVSQEPILPDCSIAENIAYGNS	1137
QY	1140	RVYSHEEIVQAKEANTHFIEITLPEKYNTRYGDKGTOLSGGQKORATARALVQPHIL	1199
DB	1138	RVYSQBEIVRAAKEANIHFIESLPNKYSTKVGDKGTOLSGGQKORATARALVQPHIL	1197
QY	1200	LLDEATSALDTESEKVVQBALDKAREGRTCIIVIAHRLSTIQNADLIVFPQNGKVEHGTH	1259
DB	1198	LLDEATSALDTESEKVVQBALDKAREGRTCIIVIAHRLSTIQNADLIVFPQNGKVEHGTH	1257
QY	1260	QQLLAOKGIYFSMVSVQAGAKR	1281
DB	1258	QQLLAOKGIYFSMVSVQAGTKR	1279
RESULT 9			
AA	81959	AAB81959 standard; protein; 1280 AA.	
ID	AA	AAB81959 standard; protein; 1280 AA.	
AC	AA	AAB81959;	
XX	AA	AAB81959;	
DT	02-JUL-2001	(first entry)	
DE	Human	MDR1.	
XX	Human;	multi-drug resistance; cholesterol uptake;	
KW	hypercholesterolaemia; hypocholesterolaemia; atherosclerosis;		
KW	coronary artery disease; cerebral vascular disease.		
XX	Homo sapiens.		
OS	WO200121762-A2.		
PN	29-MAR-2001.		
XX	22-SEP-2000; 2000WO-US26099.		
PF	23-SEP-1999; 99US-0155819.		
XX	(BARN-) BARNES-JEWISH HOSPITAL.		
PA	Stenson WF, Tessner T;		
XX	WPI; 2001-328100/34.		
DR	Modulating cellular cholesterol uptake for treating hyper or		
XX	hypercholesterolemia, involves administering an agent that inhibits or		
PT	increases the expression of multiple drug resistance-1.		
PT	Example 1; Page 43-47; 47pp; English.		
XX	The present invention describes a method of modulating cholesterol uptake		
CC	in cells, involving administering an agent capable of inhibiting the		
CC	multi-drug resistance protein MDR1. This is useful in the prevention and		
CC	treatment of cholesterol-related diseases, including		
CC	hypercholesterolaemia, hypocholesterolaemia, atherosclerosis, coronary		
CC	artery disease and cerebral vascular diseases. The present sequence is		
CC	the human MDR1 protein.		
XX	Sequence 1280 AA;		
SQ	Query Match	90.6%; Score 5866; DB 22; Length 1280;	

Best Local Similarity 90.8%; Pred. No. 0;				
Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;				
QY	1	MDPEGGRKGS-	EKNFWKMGKSKKKKKKPTVSTFAMFRYSNWLDRMLVLTGMAAI	59
DB	1	MDLEGDRNGAKKNFFKLNKS-	EKDKKKKPTSVSFMSFRYSNWLKDKLYMVGTLAAI	59
QY	60	IHGAAALPLMLVFGNMTDSFANAGISRNKTPFVLINESITNTQHF	INHLEEREMTVAYY	119
DB	60	IHGAGLPLMLVFGEMTDIFANAG-	NLEDLSNITNRSNDINDTGFMMN-LEEDMTRIAYY	117
QY	120	YSGIGAGVLVAAVYIQVSWFCLAAAGROILKIRKQFFHAIMRQEGICWFDVHDVGEINLRLTD	179	
DB	118	YSGIGAGVLVAAVYIQVSWFCLAAAGROIHKIRKQFFHAIMRQEGICWFDVHDVGEINLRLTD	177	
QY	180	DVSKINEGIDKIGMFFOSIATFTFFGTVGTRGKWLTVLILAI	SPVLGISAALWAKILS	239
DB	178	DVSKINEGIDKIGMFFOSIATFTFTGTVGTRGKWLTVLILAI	SPVLGISAALWAKILS	237
QY	240	SFTDKELLAYAKAGAVAEVLAAITRTVIAFGGQKKELERYNKNLEEA	GIGIKKAITANI	299
DB	238	SFTDKELLAYAKAGAVAEVLAAITRTVIAFGGQKKELERYNKNLEEA	RIGIKKAITANI	297
QY	300	SIGAAPELLIYASALAFWYGTSLVLSSEYTTIGQVLT	VFFSVLIGAFSIGQASPSIEAFAN	359
DB	298	SIGAAPELLIYASALAFWYGTSLVLSSEYTTIGQVLT	VFFSVLIGAFSVQASPSIEAFAN	357
QY	360	ARGAAVEIEFKIIDNKPISIDSYSKSGHKPDNKGNEFKNVHFSYPSRKEYKILKGLNLKV	419	
DB	358	ARGAAVEIEFKIIDNKPISIDSYSKSGHKPDNKGNEFKNVHFSYPSRKEYKILKGLNLKV	417	
QY	420	QSGQTVALVNGSGCGKSTTVOLMQR	LDYDPTDMGVICDQDITRINVRLHREITGVWSQEP	479
DB	418	QSGQTVALVNGSGCGKSTTVOLMQR	LDYDPTEGMVSVDQDITRINVRLREILIGVWSQEP	477
QY	480	VLFAITIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFTVLVGERGAQLSGGQKQ	539	
DB	478	VLFAITIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFTVLVGERGAQLSGGQKQ	537	
QY	540	RIATARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV	599	
DB	538	RIATARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV	597	
QY	600	IAGFDDGVIVKEGNHDELMKEGIYFKLV	TQTRGNEIELENATGESKSSDALEMSPKD	659
DB	598	IAGFDDGVIVKEGNHDELMKEGIYFKLV	TQTRGNEVELENAADESKSEIDALEMSND	657
QY	660	SGSSLIKRSTRRSIHAPOGQDRKLGTREDLNE	VPPVPSFWRILKLNSTENPYPVVGIFC	719
DB	658	SRSSLIKRSTRRSVRSQAQDRKLSTKEALDESIPPVSFWRIMKLNSTENPYPVVGIFC	717	
QY	720	AIINGGLQPAFSAIIFSRIGITRDEDEPETKRON	SFVLFLVIGIISFTFFLQGGTFF	779
DB	718	AIINGGLQPAFSAIIFSKILCVFTRIDDEPETKRON	SFVLFLVIGIISFTFFLQGGTFF	777
QY	780	KGAGEILTKRLRYMFRSMRLQDVSWFDDPKNT	TGALTTRLANDAAQVKGIGSLRLAVIT	839
DB	778	KGAGEILTKRLRYMFRSMRLQDVSWFDDPKNT	TGALTTRLANDAAQVKGIGSLRLAVIT	837
QY	840	QNIANLGTGIIISLIYQWQLTLLLAIVPI	ITAIAGVEMKMLSGQALKDKKELEGAGKIA	899
DB	838	QNIANLGTGIIISLIYQWQLTLLLAIVPI	ITAIAGVEMKMLSGQALKDKKELEGAGKIA	897
QY	900	TEAIENFRTVSLTREQKFEYMAQSLQVYRNSLRKAHIFGV	SFISITQAMMYSYAGCF	959
DB	898	TEAIENFRTVSLTREQKFEHMAQSLQVYRNSLRKAHIFGT	SFISITQAMMYSYAGCF	957
QY	960	RFGAYLVANEFMNFQDVLVLSFAIVFAGAMAVGOVSSFPADYAKAVSAAHVIMII	EKSPL	1019
DB	958	RFGAYLVVAHKLMSFEDVLVLSFAIVFAGAMAVGOVSSFPADYAKAKISAHHIIMII	EKTPL	1017
QY	1020	IDSYSPHGLKPNLTLEGNVT	FNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS	1079

Db 1018 IDSYSLEGLMPNTLEGNTVFGEWFNYPTRPDPVLPVQLGLESLVKKGQTLALVSSGCGKS 1077

QY 1080 TVQQLLERYDPLAGSLVDGKEIKHLNQWLRHLHGLIVSQBPILFDCSIAENTAYGNS 1139

Db 1078 TVQQLLERYDPLAGSLVDGKEIKHLNQWLRHLHGLIVSQBPILFDCSIAENTAYGNS 1137

QY 1140 RVVSHEEIVQAAKEANIHHFIEITPEKYNTRYGDRGTQLSGGQKQRIARALVRQPHIL 1199

Db 1138 RVVSQEEIVRAAKEANIHAFTIESLPNKYSTKVGDKGTQLSGGQKQRIARALVRQPHIL 1197

QY 1200 LDEATSALDTESEKVVQEQALDKAREGRTCVIAHRLSTIONADLIVVFQNGKVEHGT 1259

Db 1198 LDEATSALDTESEKVVQEQALDKAREGRTCVIAHRLSTIONADLIVVFQNGKVEHGT 1257

QY 1260 QQLLAQKGIYFSMVSQVQAKR 1281

Db 1258 QQLLAQKGIYFSMVSQVQAKR 1279

RESULT 10

AAB81064

ID AAB81064 standard; Protein; 1280 AA.

XX

AC AAB81064;

XX

DT 25-JUN-2001 (first entry)

XX

DE Cynomologous monkey P-glycoprotein variant 1.

XX

KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRI;

KW efflux pump.

XX

OS Macaca fascicularis.

XX

XX WO2001233565-A1.

PN

XX 05-APR-2001.

PD

XX 28-SEP-2000; 2000W0-US26592.

PF

XX 28-SEP-1999; 99US-0156921.

PR

XX 12-OCT-1999; 99US-0158818.

XX

PA (GENT-) GENTEST CORP.

XX

XX Stocker PJ, Steimel-Crespi DT, Crespi CL;

PI

XX WPT: 2001-316136/33.

DR

XX N-PSDB; AAF86127.

DR

XX

PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein

PT (PGP) and homologous PGP polypeptides are useful for predicting

PT bioavailability of compound and increasing PGP transporter activity in

PT cell -

XX

PS Claim 9; Page 57-59; 84pp; English.

XX

CC This invention relates to a polynucleotide sequence encoding a

CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the

CC PGP protein. PGP, also known as multidrug transporter, MDRI is a member

CC of the ABC transporter superfamily. The enzyme serves as an efflux pump

CC exporting small molecules across the cell membrane. The invention

CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence

CC and protein, and also that of an allelic variant. The PGP polynucleotide

CC sequence is useful for increasing PGP transporter activity in a cell.

CC Antisense sequences of the cDNA are useful for inhibiting PGP transport

CC activity in a mammalian cell. They may also be used for increasing the

CC bioavailability of a drug. The present sequence represents the

CC cynomologous monkey P-glycoprotein.

XX

SQ Sequence 1280 AA;

Best Local Similarity 90.7%; Pred. No. 0;

Matches 1163; Conservative 59; Mismatches 56; Indels 4; Gaps 4;

QY 1 MDPEGGRKGSAAE-KNFWKMGKSKKKKKKPTVSTFAMFRYSNWLDRMLVGMVMAAI 59

Db 1 MDLESGDRNGGAEEKKFFKLNKS-KDKKERKPTVSFMSFRYSNWLDRMLVGMVMAAI 59

QY 60 IHGAALPLMLVFGNMTDSFANAGISRNKPTFPVINESITNTQHTINHLEEMTYVAY 119

Db 60 IHGAGLPLMLVFGDMTDTFANAG-NLGDGLALLTSSNITDTPVPMN-LEEDMTRYAY 117

QY 120 YSGIGAGVLVAAYTOVSFWCLAGRQLTKRKOFFHAIMROEIGWDFDHDVGEINRLTD 179

Db 118 YSGIGAGVLVAAYLOVSFWCLAGRQLTKRKOFFHAIMROEIGWDFDHDVGEINRLTD 177

QY 180 DVSKINEGIDKIGMFFQSIATFTFTGIVGTRGWKLTVLILAIISPVLGLSAAWAKILS 239

Db 178 DVSKINEGIDKIGMFFQSMATFTFTGIVGTRGWKLTVLILAIISPVLGLSAAWAKILS 237

QY 240 SFTDKELLAYAKAGAAVEVLAAIRTVIARFGGOKKELERYKNKLEAKGIGIKKAITANI 299

Db 238 SFTDKELLAYAKAGAAVEVLAAIRTVIARFGGOKKELERYKNKLEAKGIGIKKAITANI 297

QY 300 SIGAAFLLIYASYALAFWYCTSLVLSSEYTGIVLTVFFSVLIGAFSGOASPSIEAFAN 359

Db 298 SIGAAFLLIYASYALAFWYCTTLVLSKEYSIGQVLTVFFSVLIGAFSGOASPSIEAFAN 357

QY 360 ARGAAVEIFKIIDNKPIDSISYKSGHKPNKGNLEFKNVHFSYPSRKEVKILKGLNLKV 419

Db 358 ARGAAVEIFKIIDNKPIDSISYKSGHKPNKGNLEFKNVHFSYPSRKEVKILKGLNLKV 417

QY 420 QSGQTVALVGNSCGKSTTVQLMORLYDPTDGHVCIDGQDIRTINVRHLREITGVVSQBP 479

Db 418 QSGQTVALVGNSCGKSTTVQLMORLYDPTDGHVCIDGQDIRTINVRHLREITGVVSQBP 477

QY 480 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNFKDTLVGPRGQALSGGQK 539

Db 478 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNFKDTLVGPRGQALSGGQK 537

QY 540 RIAIARALVRNPKILLDEATSALDTESEAVVQVVALDKARKGRTTIVIAHRLSTVRNADY 599

Db 538 RIAIARALVRNPKILLDEATSALDTESEAVVQVVALDKARKGRTTIVIAHRLSTVRNADY 597

QY 600 IAGFDGVIIVEKGNHDELMKEKGIYKVTMTQTRGNEIELENAETGESKESDALEWSPKD 659

Db 598 IAGFDGVIIVEKGNHDELMKEKGIYKVTMTQTRGNEIELENAETGESKESDALEWSPKD 657

QY 660 SGSSLIKRSTRRSIHAPOQDRKLTGKEDLNENPVPSFWIRILKLNSTEWPFYVVGIFC 719

Db 658 SGSSLIKRSTRRSVRSOGQDRKLTGKEDLNENPVPSFWIRILKLNSTEWPFYVVGIFC 717

QY 720 AIINGLOPAPSIIFRSRIIGTFRDBDPTKRONSMFSLVFLVGIISFTIFFLQGTFF 779

Db 718 AIINGLOPAPAFIIFSRIIGTFRDBDPTKRONSMFSLVFLVGIISFTIFFLQGTFF 777

QY 780 GKAGEITLTKRLRYMVFERSMLRODVSNFDDPKNTGTALTRLANDAAQVKGATGSLAVIT 839

Db 778 GKAGEITLTKRLRYMVFERSMLRODVSNFDDPKNTGTALTRLANDAAQVKGATGSLAVIT 837

QY 840 QNIANLGTGIISLIYQWLTLLLAIVPIIAIAGVEMKMLSGQALKDKKLEAGAKIA 899

Db 838 QNIANLGTGIISLIYQWLTLLLAIVPIIAIAGVEMKMLSGQALKDKKLEAGAKIA 897

QY 900 TEAIENFRIVVSLTREQKFEYMAOSLOQVPRNSLRKAHIFGVSPSITQAMWYSYACGF 959

Db 898 TEAIENFRIVVSLTREQKFEYMAOSLOQVPRNSLRKAHIFGVSPSITQAMWYSYACGF 957

QY 960 REGAYLVANEEMNFODVLLVFSAIVFGAMAVGVSSPADYAKAKVSAAHVMIIEKSP 1019

Db 958 REGAYLVANLMSFEDVLLVFSVAVVGAMAVGVSSPADYAKAKVSAAHVMIIEKSP 1017

QY 1020 IDSYPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLOGLSLEVKKGTALVVGSSGCGKS 1079

Db 1018 IDSYSBGLKPNTELEGNVTNEVFNYPTRLDIPVLQGLSLEVKKGOTLALVSSGCGKS 1077

Qy 1080 TVVOLLERFYDPLAGSVLIDGKEIKHLNVQWMLRAHLGIYSQEPILFDCSIAENIAYGDN 1139

Db 1078 TVVOLLERFYDPLAGVKVLLDGKETKQLNVQWMLRAHLGIYSQEPILFDCSISENIAYGDN 1137

Qy 1140 RVVSHESIIVQAANEANHHFIETLPEKYNTRVGDGTQLSGGQKQRIATARALVRPHIL 1199

Db 1138 RVVSEIEIVRAAKEANITHAFIESLPNKYSRVDGDKTQLSGGQKQRIATARALVRPHIL 1197

Qy 1200 LLDATSALDTESEKVVQOEALDKAREGRTCVIAHRLSTIQNADLIIVFQNGVKVKEGHT 1259

Db 1198 LLDATSALDTESEKVVQOEALDKAREGRTCVIAHRLSTIQNADLIIVFQNGVKVKEGHT 1257

Qy 1260 QLLAQRGIYFSMVSVQAGAKR 1281

Db 1258 QLLAQRGIYFSMVSVQAGAKR 1279

RESULT 11

AAB81065

ID AAB81065 standard; Protein; 1283 AA.

XX AC AAB81065;

XX DT 25-JUN-2001 (first entry)

XX DE Cynomologous monkey P-glycoprotein variant 2.

XX KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL;

XX KW efflux pump.

XX OS Macaca fascicularis.

XX FH Key Location/Qualifiers

FT Misc-difference 93..95

FT /note= "An additional 3 amino acids are present compared

XX to PGP variant AAB81064"

PN WO2001233565-A1.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26592.

XX 28-SEP-1999; 99US-0156921.

PR 12-OCT-1999; 99US-0158818.

XX (GENT-) GENTEST CORP.

XX Stocker PJ, Steimel-Crespi DT, Crespi CL;

XX WPI; 2001-316136/33.

DR N-PSDB; AAF86128.

XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein

PT (PGP) and homologous PGP polypeptides are useful for predicting

PT bioavailability of compound and increasing PGP transporter activity in

PT cell

XX Claim 9; Page 65-68; 84pp; English.

XX This invention relates to a polynucleotide sequence encoding a

CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the

CC PGP protein. MGP, also known as multidrug transporter, MDRL is a member

CC of the ABC transporter superfamily. The enzyme serves as an efflux pump

CC exporting small molecules across the cell membrane. The invention

CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence

CC and protein, and also that of an allelic variant. The PGP polynucleotide

CC sequence is useful for increasing PGP transporter activity in a cell.

CC Antisense sequences of the cDNA are useful for inhibiting PGP transport

CC activity in a mammalian cell. They may also be used for increasing the

CC bioavailability of a drug. The present sequence represents the

CC cynomologous monkey P-glycoprotein variant 2. The protein has an

CC additional 3 amino acids when compared to PGP variant 1 (AAB81065).

XX Sequence 1283 AA;

SQ Query Match 90.5%; Score 5859.5; DB 22; Length 1283;

Best Local Similarity 90.4%; Pred. No. 0;

Matches 1165; Conservative 57; Mismatches 52; Indels 15; Gaps 5;

Qy 1 MDPEGRGKSAE-KNFWKMGKKKKKKKPTVSTFAMFYSNWLDRMLVGLTMAAI 59

Db 1 MDLEGDRNGABKKFFKLNKNS-KDKRKRPVTSVSMFYSNWLDRMLVGLTMAAI 59

Qy 60 IHGAALPLMLVFGNNTDSFANAG-----ISRNKTPFVINESITNQHFINHLEEE 112

Db 60 IHGAGLPMLVFGDWTDTFANAGNLGDLGALLFNNT-----NSSNITDTPVVMN-LEED 113

Qy 113 MTTAYYYSGIGAGVLVAAYIQVSWFCLAAAGROILKIRKQFFHAIMROEIGWFDVHDVE 172

Db 114 MTRYAYYYSGIGAGVLVAAYIQVSWFCLAAAGROIHKIRKQFFHAIMROEIGWFDVHDVE 173

Qy 173 LNTRLTDDVSKINEGIGDKIGMFFOSIATFTFTGTVGTRGWKLTVLVILAIISPVLGLSAA 232

Db 174 LNTRLTDDVSKINEGIGDKIGMFFOSMATFTFTGTVGTRGWKLTVLVILAIISPVLGLSAA 233

Qy 233 IWAKILSSFTDKELLAYAKAGAVAEVLAAIRTVIAFGGOKKELERYKNKLEEKIGIK 292

Db 234 VWAKILSSFTDKELLAYAKAGAVAEVLAAIRTVIAFGGOKKELERYKNKLEEKIRIGIK 293

Qy 293 KAITANISGAFLIIYASYALAFWYGTSLVSSVETIGOVLTVFVSLVIGAFSGQASP 352

Db 294 KAITANISGAFLIIYASYALAFWYGTTLVLSKEYSIGQVLTVFVSLVIGAFSGQASP 353

Qy 353 STEAFANARGAAYEIFKIIDNKPSIDSYSGHKKPDNKGKLEFNKFNHVSYSRKEVIL 412

Db 354 STEAFANARGAAYEIFKIIDNKPSIDSYSGHKKPDNKGKLEFNKFNHVSYSRKEVIL 413

Qy 413 KGLNLKVSQGTVALVGNSSGCKSTTVQLMQRLYDPTDGMVCDIGODIETINVRHLREIT 472

Db 414 KGLNLKVSQGTVALVGNSSGCKSTTVQLMQRLYDPTDGMVSDVDGODIETINVRHLREIT 473

Qy 473 GVVSQEPVLFPATTIAENIRYGRNVTMDTEKAVKCANAYDFIMKLPNKFDTLVGERGAQ 532

Db 474 GVVSQEPVLFPATTIAENIRYGRNVTMDTEKAVKCANAYDFIMKLPNKFDTLVGERGAQ 533

Qy 533 LSGGQKQRIATARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLS 592

Db 534 LSGGQKQRIATARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLS 593

Qy 593 TVRNADVIAFGDDGVIVEKGNHDELMKKEGIYFKLVTMTGRNEIELENATGESKESDA 652

Db 594 TVRNADVIAFGDDGVIVEKGNHDELMKKEGIYFKLVTMTGRNEIELENATGESKESDA 653

Qy 653 LEMSPKDSGSLIKRRSTRRSIHAPQGRKLTGKEDLNENPPVPSFWRLKLNSTEWPY 712

Db 654 LEMSHDSDGSLIKRRSTRRSVRSVRSQGRKLTGKEDLNENPPVPSFWRLKLNSTEWPY 713

Qy 713 FVVGFCALINGGLQAPAFSIFRSIIGITFRDEDPETKQNSNMFSVLVLVIGIISFIF 772

Db 714 FVVGFCALINGGLQAPAFSIFRSIIGITFRDDEATKQNSNLFSLLFLVLVIGIISFIF 773

Qy 773 FLOGFTFGKAGEILTKRLRYMVFMSRLQDVSWFDDPKNTTGALTTRLANDAAQVKAIG 832

Db 774 FLOGFTFGKAGEILTKRLRYMVFMSRLQDVSWFDDPKNTTGALTTRLANDAAQVKAIG 833

Qy 833 SRLAVITQNIANTLGTGIIISLIYQWQLTLLLAIVPIIAITAGVVENKMLSGQALKDKKEL 892

Db 834 SRLAVITQNIANTLGTGIIISLIYQWQLTLLLAIVPIIAITAGVVENKMLSGQALKDKKEL 893

Qy 893 EGAGKIATEAIEINFRVTVSVSTRQKPEYMTAQSLQVPYRNSLRKAHIFGVSFSITQAMMY 952

Db 894 EGAGKIATEAIEINFRVTVSVSTRQKPEYMTAQSLQVPYRNSLRKAHIFGVSFSITQAMMY 953

QY 953 FSYACGFRGAYLVANEFMNFQDVLVLSAIVFGAMAVGVQVSSFPADYAKAKVSAAHVIM 1012
 Db 954 FSYACGFRGAYLVANEFMNFQDVLVLSAIVFGAMAVGVQVSSFPADYAKAKVSAAHVIM 1013
 QY 1013 IIEKSPILDSYSPGLKXNTLEGNTFNEVFNTPTRPDIPVLOGLSLEVKKGTALVG 1072
 Db 1014 IIEKSPILDSYSPGLKXNTLEGNTFNEVFNTPTRPDIPVLOGLSLEVKKGTALVG 1073
 QY 1073 SSGGKSTVWQLLRFYDPLAGSVLIDGKEIKLNWOLRAHLGIVSQEPTLPDCSTAE 1132
 Db 1074 SSGGKSTVWQLLRFYDPLAGSVLIDGKEIKLNWOLRAHLGIVSQEPTLPDCSTAE 1133
 QY 1133 IAYGNSRVSHBEIVQAAKANIHHFLETPEKYNTRVGDGKQLSGGQKQRIARAL 1192
 Db 1134 IAYGNSRVSHBEIVQAAKANIHHFLETPEKYNTRVGDGKQLSGGQKQRIARAL 1193
 QY 1193 VRQPHILLDEATSEKVVQVQALDKAREGTCIVTAHRLSTIONADLIIVFQNGK 1252
 Db 1194 VRQPHILLDEATSEKVVQVQALDKAREGTCIVTAHRLSTIONADLIIVFQNGR 1253
 QY 1253 VKEHGTQOOLLAQKGIYFSMVSVOGAKR 1281
 Db 1254 VKEHGTQOOLLAQKGIYFSMVSVOGAKR 1282

RESULT 12
 AAY58187
 ID AAY58187 standard; Protein; 1280 AA.
 XX AAY58187;
 XX DT 14-MAR-2000 (first entry)
 XX DE Human G185V mutant multidrug resistance-1 (MDR-1) protein.
 XX KW Multidrug resistance; MDR-1; P-glycoprotein;
 KW transmembrane efflux pump; haematopoietic stem cell; transduction;
 KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
 KW gene therapy; gene replacement; genetic defect; thalassaemia;
 KW Gaucher's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
 KW cytokine; mutant; mutein.
 XX OS Homo sapiens.
 OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Misc-difference 185 /note= "this residue is Gly in the wild-type MDR-1
 (AAY58186)"
 FT WO9961589-A2.
 XX PD 02-DEC-1999.
 XX PF 27-MAY-1999; 99MO-US11825.
 XX PR 28-MAY-1998; 98US-0086988.
 XX PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 XX PI Sorrentino B, Bunting K;
 XX WPI: 2000-072615/06.
 DR N-PSDB; AAZ49333.
 XX Ex vivo expansion of haematopoietic stem cells transduced with a
 PT sequence encoding human multidrug resistance-1, used for bone marrow
 PT transplantation -
 XX Example 1; Page 82-90; 113pp; English.
 XX This sequence represents human G185V mutant multidrug resistance
 CC protein MDR-1, where the Gly residue at position 185

CC of the wild-type protein (AAY58186) is replaced by Val. MDR-1 is a
 CC transmembrane efflux pump, responsible for the export of drugs from
 CC cells, particularly cancer cells. The wild-type MDR-1 shows increased
 CC resistance to etoposide and decreased resistance to vinca alkaloids
 CC compared with the G185V mutant. The invention relates to transducing
 CC haematopoietic stem cells with nucleic acid encoding an MDR protein
 CC and culturing the modified cells. The modified haematopoietic stem
 CC cells are useful in bone marrow transplantation (to reconstitute
 CC haematopoietic systems in patients who have undergone chemotherapy or
 CC radiation therapy) and in ex vivo gene therapy of genetic defects in
 CC cells derived from haematopoietic stem cells, e.g., thalassaemia,
 CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
 CC cells can also be used to identify factors involved in regulating
 CC proliferation and differentiation in haematopoietic stem cells.
 CC Haematopoietic stem cells that express MDR-1 will be protected against
 CC chemotherapeutic agents, so can be engrafted while the patient is
 CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
 CC provides sufficient cells to permit standard biochemical analysis.
 CC Overexpression of MDR-1 allows cytokine-driven expansion of
 CC haematopoietic stem cells by at least 10-fold compared with a maximum
 CC of 4-fold in known procedures.
 XX Sequence 1280 AA;
 SQ
 Query Match 90.5%; Score 5857; DB 21; Length 1280;
 Best Local Similarity 90.7%; Pred. No. 0;
 Matches 1163; Conservative 55; Mismatches 60; Indels 4; Gaps 4;
 QY 1 MDPEGGKGSAA-EKNFWMKGSKKKKKKPTVTFTAFMFRYSNWLRLYMLVGTMAAI 59
 Db 1 MDLEGRNGGAKKKNFKLNKS-EKDKKKPTVSVFMSFRYSNWLRLYMLVGTMAAI 59
 QY 60 IHGAALPLMLVFGNMTDSFANAGISRNKTPFVINESITNNQTHFINLEEMTYAY 119
 Db 60 IHGAGLPLMLVFGNMTDSFANAG-NLEDLSNITNRSDINDTGFEMN-LEEDMTRYAY 117
 QY 120 YSGIGAGVLVAAYIQVSWFCLAAAGROILAKIRQFHAIIMRQETGDFVHDVGLNTRLD 179
 Db 118 YSGIGAGVLVAAYIQVSWFCLAAAGROILAKIRQFHAIIMRQETGDFVHDVGLNTRLD 177
 QY 180 DVSKINEGIDGKGMFFQSIAITFTTGFVGTGRWKLTILVLAISPVLGSLAAIAWAKILS 239
 Db 178 DVSKINEVIGDKIGMFFQSMATFTTGFVGTGRWKLTILVLAISPVLGSLAAIAWAKILS 237
 QY 240 SFTDKELLAYAKAGAAVEEVLAAIRTVIAFGGQKKELELYNKNLEAKGIGIKKAITANI 299
 Db 238 SFTDKELLAYAKAGAAVEEVLAAIRTVIAFGGQKKELELYNKNLEAKGIGIKKAITANI 297
 QY 300 SIGAFLLIYASYALAFWGTGTLVLSGEYSIGQVLTFFSVLIGAFSGQASPSFEAFAN 359
 Db 298 SIGAFLLIYASYALAFWGTGTLVLSGEYSIGQVLTFFSVLIGAFSGQASPSFEAFAN 357
 QY 360 ARGAAVEIFKIIDNKPSIDSYSGSKGHKPDNIKGNLEFNHVSFYSRKEVKILGLNLKV 419
 Db 358 ARGAAVEIFKIIDNKPSIDSYSGSKGHKPDNIKGNLEFNHVSFYSRKEVKILGLNLKV 417
 QY 420 QSGQTVLVNCGSGCKSTTVQLMQLYDPTDGMWIDQDITINVRHLREITGVVSE 479
 Db 418 QSGQTVLVNCGSGCKSTTVQLMQLYDPTDGMWIDQDITINVRHLREITGVVSE 477
 QY 480 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAOLSGGQK 539
 Db 478 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAOLSGGQK 537
 QY 540 RIATARALVRNPKILLDEATSEAVVQVVALDKARKGRITIVIAHRLSTVRNADV 599
 Db 538 RIATARALVRNPKILLDEATSEAVVQVVALDKARKGRITIVIAHRLSTVRNADV 597
 QY 600 IAGFDGIVIVEKGNHDELMKEGIYFKLVMTQTRONELENAAGESKESDALEMSPKD 659
 Db 598 IAGFDGIVIVEKGNHDELMKEGIYFKLVMTQTRONELENAAGESKESDALEMSPKD 657
 QY 660 SSSSLIKRRSTRSITHAPQGDQRKLTGKEDLNENVPVPSFWRLKLNSTWPFVVGIFC 719

Db 658 SRSLRKSTRSRVSGSAQDRKSLKREALDESIPPSFWRMKLNTEWPFVVGVC 717
QY 720 AIINGLOPAFISIFRSIIGTIFRDEDPETKRONSMFSLVFLVIGIISFTIFFLOQFTF 779
Db 718 AIINGLOPAFAIFISKIIGVTRIDDPETKRONSLFLSLFLALGIISFTIFFLOQFTF 777
QY 780 GRAGEILTLRLRYMFRSMLRDVSWFDDPKNTTGALTTRLANDAQAQVKGAGSLAVIT 839
Db 778 GRAGEILTLRLRYMFRSMLRDVSWFDDPKNTTGALTTRLANDAQAQVKGAGSLAVIT 837
QY 840 QNIANLGTGIIISLIYGMQLTLLLAIVPIIAIAGVEMKMLSGQALKDKELEGAGKIA 899
Db 838 QNIANLGTGIIISFIYGMQLTLLLAIVPIIAIAGVEMKMLSGQALKDKELEGAGKIA 897
QY 900 TEAIENFRVWSLTRQKFEYMYAQSLOVPYRNSLRKAHIFGVFSITQAMMYFYAGCF 959
Db 898 TEAIENFRVWSLTORQKFEYMYAQSLOVPYRNSLRKAHIFGITFSFTQAMMYFYAGCF 957
QY 960 RFGAYLVANEFNFQDVLVFSALVFGAMAGVSSFPADYAKAKVSAHVIMITEKSP 1019
Db 958 RFGAYLVANHLKMSFEDVLVFSAVFGAMAGVSSFPADYAKAKISRAHIIMITEKTP 1017
QY 1020 IDSYSPLGLKPNLTLEGNTFNEVFNYPTRPDIPVLQGLSLVKKGQTLALVSGSGCGKS 1079
Db 1018 IDSYSLEGLMPNLTLEGNTFGEVFNYPTRPDIPVLQGLSLVKKGQTLALVSGSGCGKS 1077
QY 1080 TVVQLLERYFDPLAGSVLDGKEIKHLNQWLRAHLGIVSQBPILFDCSIAENIAYGDS 1139
Db 1078 TVVQLLERYFDPLAGSVLDGKEIKHLNQWLRAHLGIVSQBPILFDCSIAENIAYGDS 1137
QY 1140 RVYSHEIEVQAKEANHHFIETLPEKYNTRYGDKTQSLGGQKORIALARALVRQPHIL 1199
Db 1138 RVYSQEIIVRAKEANHHAFIESLPNKYSTKYGDKTQSLGGQKORIALARALVRQPHIL 1197
QY 1200 LLDDEATSDTESEKVVQALDKAREGRTCVIAHRLSTIQNADLIVFPQNGKVEHGH 1259
Db 1198 LLDDEATSDTESEKVVQALDKAREGRTCVIAHRLSTIQNADLIVFPQNGKVEHGH 1257
QY 1260 QLLAOKGIYFMSVQAGAKR 1281
Db 1258 QLLAOKGIYFMSVQAGTKR 1279

RESULT 13
AAB81066
ID AAB81066 standard; Protein; 1280 AA.
XX AAB81066;
AC
XX
DT 25-JUN-2001 (first entry)
XX
DE Human P-glycoprotein SEQ ID 5.
XX
KW Cynologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL;
KW efflux pump; human.
XX
OS Homo sapiens.
XX
PN W0200123565-A1.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26592.
XX
PR 28-SEP-1999; 99US-0156921.
PR 12-OCT-1999; 99US-0158818.
XX
PA (GENT-) GENTEST CORP.
XX
XX
PI Stocker PJ, Steimel-Crespi DT, Crespi CL;
XX WPI; 2001-316136/33.

XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
cell -
XX
XX
PS Claim 6; Page 68-71; 84pp; English.
XX
XX This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDRL is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents a human
CC P-glycoprotein.
XX
XX Sequence 1280 AA;
SQ

Query Match 90.5%; Score 5857; DB 22; Length 1280;
Best Local Similarity 90.7%; Pred. No. 0;
Matches 1163; Conservative 55; Mismatches 60; Indels 4; Gaps 4;

QY 1 MDPEGGRGSA-EKNFWMKMGKKKKKKPTVSTFAMFRYSNWLDRMLYMLVGTMAAI 59
Db 1 MDLEDRNGGAKKNFFKLNKS-EKDKKPKVSVFSEMYSNWLDRMLYMLVGTMAAI 59

QY 60 IHGAALPLMLVFGNMTDSFANAGISRNKTPPVITINSITNTTOHFNHLEEMTTVAY 119
Db 60 IHGAGLPLMLVFGEMTDIFANAG-NLEDLSNITNRSNDINDTGFNMN-LEEDMTRYAY 117

QY 120 YSGIGAGVLVAAVTOVSFWCLAAAGROILKIRKQFFHAIMRQEIFGDFVHDVGEINLTLD 179
Db 118 YSGIGAGVLVAAVTOVSFWCLAAAGROILKIRKQFFHAIMRQEIFGDFVHDVGEINLTLD 177

QY 180 DVSKINEGIGDKIGMFFQSIATFTFTGTVGTGKMLTILVILAIISPVGLSAAIAWAKILS 239
Db 178 DVSKINEVIGDKIGMFFQSMATFTFTGTVGTGKMLTILVILAIISPVGLSAAIAWAKILS 237

QY 240 SFTDKELLAYAKAGAAVEVLAAITVIAFGGQKKELEKYNKLEAKGIGIKKAITANI 299
Db 238 SFTDKELLAYAKAGAAVEVLAAITVIAFGGQKKELEKYNKLEAKGIGIKKAITANI 297

QY 300 STGAFLLIYASALAFWYGTSLVLSSEYITGQVLTIVFFSVLIGAFSIGQASPSIEAFAN 359
Db 298 STGAFLLIYASALAFWYGTSLVLSSEYITGQVLTIVFFSVLIGAFSIGQASPSIEAFAN 357

QY 360 ARGAAEYIFKIIDNKPSIDSYSGSHKPDNIKGNLEFKNHFSYPSRKEVKILKGLNLKV 419
Db 358 ARGAAEYIFKIIDNKPSIDSYSGSHKPDNIKGNLEFKNHFSYPSRKEVKILKGLNLKV 417

QY 420 OSGQTVLVGNSGCKSTTVQLMQRLYDPTGMCVDCQDITINVRHLRITGVVSOEP 479
Db 418 OSGQTVLVGNSGCKSTTVQLMQRLYDPTGMCVDCQDITINVRHLRITGVVSOEP 477

QY 480 VLFATTIAENTRYGRENVTMDIEKAVKEANAYDFIMKLPNKFDITLVGERGAQLSGGQK 539
Db 478 VLFATTIAENTRYGRENVTMDIEKAVKEANAYDFIMKLPNKFDITLVGERGAQLSGGQK 537

QY 540 RIATARALVRNPKILLDLDEATSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADV 599
Db 538 RIATARALVRNPKILLDLDEATSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADV 597

QY 600 IAGFDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNETELENADESKEIDALEMSPKD 659
Db 598 IAGFDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNETELENADESKEIDALEMSND 657

QY 660 SGSSLIKRRSTRRSIHAPQGDQRKLTGKEDLNENPPYSFWRILKLNSTENPFFVVGIFC 719
Db 658 SGSSLIKRRSTRRSIHAPQGDQRKLTGKEDLNENPPYSFWRILKLNSTENPFFVVGIFC 717

Db 658 SRSSLIKRSTRSRVRSQAQDRKLSTKEALDESIPVPSFWRIMKLNLTENPYPVWGVC 717
 Qy 720 AILINGLOPAFISIIISRIIGTTRDEDEPETKRONSNFSLFLVLGIISFTIFFLOQFTF 779
 Db 718 AILINGLOPAFIIISFKIIGVTRIDDPETKRONSNFSLFLALGIIISFTIFFLOQFTF 777
 Qy 780 GKAGEIITKRLRYMVFRSMLRQDYSWFDDPKNTTGALTTRLANDAAOVKGAIGSRLAVIT 839
 Db 778 GKAGEIITKRLRYMVFRSMLRQDYSWFDDPKNTTGALTTRLANDAAOVKGAIGSRLAVIT 837
 Qy 840 QNIANLGTGIIISIIYQWQLTLLLAIVPIIAIAGVEMKMLSGQALKDKKELEGACKIA 899
 Db 838 QNIANLGTGIIISIIYQWQLTLLLAIVPIIAIAGVEMKMLSGQALKDKKELEGACKIA 897
 Qy 900 TEATENERTVSLTREOKFEYMAQSLQVYRNSLRKAHIFGVYSFSTQAMWYFSYAGCF 959
 Db 898 TEATENERTVSLTREOKFEYMAQSLQVYRNSLRKAHIFGVYSFSTQAMWYFSYAGCF 957
 Qy 960 RFGAYLVANFPMFQDVLVFSIAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSP 1019
 Db 958 RFGAYLVANFPMFQDVLVFSIAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSP 1017
 Qy 1020 IDSYSPHCLPKNTLEGNNVTNEVFNYPTRPDIPVLOGLSLEVKKGOTLALVSGSGCKS 1079
 Db 1018 IDSYSPHCLPKNTLEGNNVTNEVFNYPTRPDIPVLOGLSLEVKKGOTLALVSGSGCKS 1077
 Qy 1080 TVOLLERFYDPLAGSVLDGKEIKHLNVQWLAHILGIVSQEPILFPCSAENIAYGDNS 1139
 Db 1078 TVOLLERFYDPLAGSVLDGKEIKHLNVQWLAHILGIVSQEPILFPCSAENIAYGDNS 1137
 Qy 1140 RVVSHIEIVQAAKANTHFIETLPEKYNTRVGDGKQQLSGGQKQRIATARALVRPHIL 1199
 Db 1138 RVVSHIEIVQAAKANTHFIETLPEKYNTRVGDGKQQLSGGQKQRIATARALVRPHIL 1197
 Qy 1200 LLDGATLSDTSEKVVQVQALDKAREGTCIVIAHRLSTIQNADLIIVFONGVKKEHGH 1259
 Db 1198 LLDGATLSDTSEKVVQVQALDKAREGTCIVIAHRLSTIQNADLIIVFONGVKKEHGH 1257
 Qy 1260 QQLLAQGIYFSMVSVQAGAKR 1281
 Db 1258 QQLLAQGIYFSMVSVQAGTKR 1279

RESULT 14
 AAEE00306
 ID AAEE00306 standard; Protein; 1280 AA.
 AC AAEE00306;
 XX AAEE00306;
 DT 13-JUN-2001 (first entry)
 XX Human P-glycoprotein (PGP) #1.
 DE Human; P-glycoprotein; PGP; multidrug transporter; MDRL;
 KW drug bioavailability; transgenic animal; genetic model.
 XX Homo sapiens.
 OS WO200123540-A2.
 XX 05-APR-2001.
 XX 28-SEP-2000; 2000WO-US26767.
 PF 28-SEP-1999; 99US-0156510.
 XX (GENT-) GENTEST CORP.
 PA Stocker PJ, Steelmel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 PI WPI; 2001-235373/24.
 XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful

PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors -
 XX Claim 16; Page 78-80; 11pp; English.
 CC The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDRL) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is human P-glycoprotein (PGP). The
 CC PGP enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.
 XX Sequence 1280 AA;
 SQ Query Match 90.5%; Score 5857; DB 22; Length 1280;
 Best Local Similarity 90.7%; Pred. No. 0;
 Matches 1163; Conservative 55; Mismatches 60; Indels 4; Gaps 4;
 QY 1 MDPEGGRKGS-A-EKNFMKMGKSKKKEKPTVTSTFAMFRYSNMDLRYMLVGTMAAI 59
 Db 1 MDLEGRNGGAKKFNKFNKNS-EKDKKEKPTVSTFAMFRYSNMDLRYMLVGTMAAI 59
 QY 60 IHGAALPLMLVFGNMTDSFANAGISRNKTPFVILINESITNTQHFHLEBEWYAY 119
 Db 60 IHGAGLPLMLVFGNMTDSFANAGISRNKTPFVILINESITNTQHFHLEBEWYAY 117
 QY 120 YSGIGAGVLVAAYIQVSWCLAGRQILKIRKQFFHATMRQIGWDFVHDVDELNTRLD 179
 Db 118 YSGIGAGVLVAAYIQVSWCLAGRQILKIRKQFFHATMRQIGWDFVHDVDELNTRLD 177
 QY 180 DYSKINEGIGDKIGMFFQSIATFTTGFVGTGRWKLTVILAISPVLGSLAAIWAKILS 239
 Db 178 DYSKINEGIGDKIGMFFQSIATFTTGFVGTGRWKLTVILAISPVLGSLAAIWAKILS 237
 QY 240 SFTDKELLAYAGAVAEVLAARTVIAFGQKKELERYNNKLEAKGIGIKKAITANI 299
 Db 238 SFTDKELLAYAGAVAEVLAARTVIAFGQKKELERYNNKLEAKGIGIKKAITANI 297
 QY 300 SIGAFLLIYASYALAFWYGTSLVLSSEVTIGOVLTFFSVLIGAFSIGQASPSFEAFAN 359
 Db 298 SIGAFLLIYASYALAFWYGTSLVLSSEVTIGOVLTFFSVLIGAFSIGQASPSFEAFAN 357
 QY 360 ARGAAEYIFKIIDNKPSIDSYSGHKKPDNKNLEFNHVSFYSRKEVKILGLNLKV 419
 Db 358 ARGAAEYIFKIIDNKPSIDSYSGHKKPDNKNLEFNHVSFYSRKEVKILGLNLKV 417
 QY 420 QSGQTVLVGNSGCGKSTTVQMLMORLYDPTDGMWICDQDITINVRHLREITGVVSE 479
 Db 418 QSGQTVLVGNSGCGKSTTVQMLMORLYDPTDGMWICDQDITINVRHLREITGVVSE 477
 QY 480 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGEGAQLSGGQK 539
 Db 478 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGEGAQLSGGQK 537
 QY 540 RIATARALVRNPKILLDLDEATSAIDTSEAVVQVVALDKARKGRTTIVIAHRLSTVRNAD 599
 Db 538 RIATARALVRNPKILLDLDEATSAIDTSEAVVQVVALDKARKGRTTIVIAHRLSTVRNAD 597
 QY 600 TAGFDGIVVEKGNHDELMKEGIFYFLVMTQTRGNETELENATGESKSDALEMSPKD 659
 Db 598 TAGFDGIVVEKGNHDELMKEGIFYFLVMTQTRGNETELENATGESKSDALEMSPKD 657

QY 660 SGSSLLKRSTRSRSHAPQODRKLGTKEIDLNNVPPVFWIRLKLNSTENPYPYFVVGIFC 719
Db 658 SRSSLLKRSTRSRVSGSQADRKLTKEALDESIPPVFWIRKLNLNTEWYPYFVVGIFC 717
QY 720 AINGLQPAFAIISRIISIGITFTRDEDETKRQNSMFSVLVLGIISFTFFLQGTFF 779
Db 718 AINGLQPAFAIISRIISIGITFTRDEDETKRQNSLFLALGIIISFTFFLQGTFF 777
QY 780 GRAGETLTKRLMYFRSMLRQDVSWFDDPKNTTGALTRLANDAAQVKGAGISRLAVIT 839
Db 778 GRAGEILLTKRLMYFRSMLRQDVSWFDDPKNTTGALTRLANDAAQVKGAGISRLAVIT 837
QY 840 QNIANLGTGIIISLIYGWQLTLLLLAIVPIIAIAGVVKMMLSGQALKDKELEGAGKIA 899
Db 838 QNIANLGTGIIISFIYGWQLTLLLLAIVPIIAIAGVVKMMLSGQALKDKELEGAGKIA 897
QY 900 TEAIENFRVVSILTRQKEIYMYAQSLQVPYRNSLRKKAHIFGVFSITQAMMYFYAGCF 959
Db 898 TEAIENFRVVSILTRQKEIYMYAQSLQVPYRNSLRKKAHIFGITFSFTQAMMYFYAGCF 957
QY 960 RFGAYLVANEFMNFQDVLVLSAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPL 1019
Db 958 RFGAYLVANLKSFEDEVLLVFSNVFGAMAVGVSSFPADYAKAKISAHHIIMIEKTPL 1017
QY 1020 IDSYPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS 1079
Db 1018 IDSYSTGLMPNLTLEGNVTFGEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS 1077
QY 1080 TVVOLLERYDPLAGSVLIDGKEIKHLNVQWLRHAHIGIVSQPIILFDCSIAENIAYGNS 1139
Db 1078 TVVOLLERYDPLAGSVLIDGKEIKHLNVQWLRHAHIGIVSQPIILFDCSIAENIAYGNS 1137
QY 1140 RVYSHEEIVQAAKEANIHHFIETLPKYNTVRVGDGKTQLSGGQKORIAIARALVRQPHIL 1199
Db 1138 RVYSQEIIVRAKEANIHAFIESLPNKYSTKVGDKGTQLSGGQKORIAIARALVRQPHIL 1197
QY 1200 LLDATSALDTESEKVVQBALDKAREGRTCIIVAHRLSTIQNADLIVFQNGKVKHEGTH 1259
Db 1198 LLDATSALDTESEKVVQBALDKAREGRTCIIVAHRLSTIQNADLIVFQNGKVKHEGTH 1257
QY 1260 QLLAOKGIYFVSMVSVQAGAKR 1281
Db 1258 QLLAOKGIYFVSMVSVQAGTKR 1279
RESULT 15
AAP70452
ID AAP70452 standard; Protein; 1280 AA.
AC AAP70452;
XX AAP70452;
XX 21-MAY-1991 (first entry)
DT XX
XX Sequence encoded by human multi-drug resistance-1 (mdr1) cDNA
DE from clones lambda HDRL0.5 and 104.
DE XX
XX Chemo-therapy resistant tumour cell; P-glycoprotein.
KW Homo sapiens.
XX W08705943-A.
XX 08-OCT-1987.
XX 26-MAR-1987; 87WO-US00758.
XX 01-AUG-1986; 86US-0892575.
PR 28-MAR-1986; 86US-0845610.
XX (UNII) UNIV OF ILLINOIS.
XX Roninson IB, Pastan IH, Gottesman MM;
XX

DR WPI; 1987-291656/41.
DR N-PSDB; AAN70752.
XX
PT DNA for multi-drug resistance in human cells - used to detect
PT chemotherapy-resistant tumour cells and for producing
PT polypeptide(s) for diagnosis and therapy
XX
PS Claim 4(a); Table 5, pp30-39; 61pp; English.
XX
CC The human multi-drug resistant KB carcinoma cell lines were used as
CC the source of the mdr1 gene nucleic acid sequences (AAN70751). To
CC obtain cDNA clones of the mdr1 gene (AAN70752), poly (A) and RNA was
CC used. Analysis of the AA sequence presented in (AAN70752) indicates
CC that the mdr1 gene product is likely to be a transmembrane protein.
CC The presence of transmembrane domains and potential glycosylation
CC sites is consistent with the mdr1 protein being related to the
CC P-glycoprotein.
XX
SQ Sequence 1280 AA;
Query Match 90.4%; Score 5854; DB 8; Length 1280;
Best Local Similarity 90.6%; Pred. No. 0;
Matches 1162; Conservative 56; Mismatches 60; Indels 4; Gaps 4;
QY 1 MDPEGGRKGS-A-EKNFWKMGKKKKKKPTVSTFAMFRYSNWLDRLYMLVGTMAAI 59
Db 1 MDLEGDRNGAKKNFTKLNKS-EKDKKAKKPTVSVFMSFRYSNWLDKLYMVVGTIAAI 59
QY 60 IHGAALPLMLVFGNMTDSFANAGISRNTFFPVIINESITNTQHTFHNHLEEMTYY 119
Db 60 IHGAGLPLMLVFGEMTDIFANAG-NLEDLSNITNRSNDINDTGFWMN-LEEDMTRVAY 117
QY 120 YSGIGAGVLVAAVYQVSWFCLAGRQILKIRKQFFHAIMRQEGWFDVHDVGEINTRLTD 179
Db 118 YSGIGAGVLVAAVYQVSWFCLAGRQIHKIRKQFFHAIMRQEGWFDVHDVGEINTRLTD 177
QY 180 DVSKINEGIGDKIGMFQSIATFTFTGTVGRGKTLTLVLAISPVGLSAAIWAAIL 239
Db 178 DVSKINQVIGDKIGMFQSMATFTFTGTVGRGKTLTLVLAISPVGLSAAVWAAIL 237
QY 240 SFTDKELIYAKAGAAVEVLAAITRTVIAFGGQKKELEKYNKNLEAKGIGIKKAITANI 299
Db 238 SFTDKELIYAKAGAAVEVLAAITRTVIAFGGQKKELEKYNKNLEAKGIGIKKAITANI 297
QY 300 SIGAAFLIIYASALAFWYGTSLVSLEYTIGQVLTVEFSVLIGAFSTIGQASPSIEAFAN 359
Db 298 SIGAAFLIIYASALAFWYGTTLVLSGEYSIGQVLTVEFSVLIGAFSVGQASPSIEAFAN 357
QY 360 ARGAAEYIFKIIDNKPISDSYKSGHKPDNIKGNLEKFNHVSYPSPRKEYKILKGLNKV 419
Db 358 ARGAAEYIFKIIDNKPISDSYKSGHKPDNIKGNLEKFNHVSYPSPRKEYKILKGLNKV 417
QY 420 QSGQTVALLVNGSGCGKSTTVQIMQRLYDPTDMVCIDGQDITRTINVRHLREITGVVSOEP 479
Db 418 QSGQTVALLVNGSGCGKSTTVQIMQRLYDPTDMVCIDGQDITRTINVRHLREITGVVSOEP 477
QY 480 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPHKFTDLVGERGAQLSGGQK 539
Db 478 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPHKFTDLVGERGAQLSGGQK 537
QY 540 RIATARALVRNPKILLDDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 599
Db 538 RIATARALVRNPKILLDDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 597
QY 600 IAGFDGIVVEKGNHDELMKEGIFKLVMTQTRGNELEENATGESKESDALEMSPKD 659
Db 598 IAGFDGIVVEKGNHDELMKEGIFKLVMTQTRGNELEENATGESKESDALEMSND 657
QY 660 SCSSLLIKRSTRSRSHAPQODRKLGTKEIDLNNVPPVFWIRLKLNSTENPYPYFVVGIFC 719
Db 658 SRSSLLIKRSTRSRVSGSQADRKLTKEALDESIPPVFWIRKLNLNTEWYPYFVVGIFC 717
QY 720 AINGLQPAFAIISRIISIGITFTRDEDETKRQNSMFSVLVLGIISFTFFLQGTFF 779
Db 718 AINGLQPAFAIISRIISIGITFTRDEDETKRQNSLFLALGIIISFTFFLQGTFF 777
QY 780 GRAGETLTKRLMYFRSMLRQDVSWFDDPKNTTGALTRLANDAAQVKGAGISRLAVIT 839
Db 778 GRAGEILLTKRLMYFRSMLRQDVSWFDDPKNTTGALTRLANDAAQVKGAGISRLAVIT 837
QY 840 QNIANLGTGIIISLIYGWQLTLLLLAIVPIIAIAGVVKMMLSGQALKDKELEGAGKIA 899
Db 838 QNIANLGTGIIISFIYGWQLTLLLLAIVPIIAIAGVVKMMLSGQALKDKELEGAGKIA 897
QY 900 TEAIENFRVVSILTRQKEIYMYAQSLQVPYRNSLRKKAHIFGVFSITQAMMYFYAGCF 959
Db 898 TEAIENFRVVSILTRQKEIYMYAQSLQVPYRNSLRKKAHIFGITFSFTQAMMYFYAGCF 957
QY 960 RFGAYLVANEFMNFQDVLVLSAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPL 1019
Db 958 RFGAYLVANLKSFEDEVLLVFSNVFGAMAVGVSSFPADYAKAKISAHHIIMIEKTPL 1017
QY 1020 IDSYPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS 1079
Db 1018 IDSYSTGLMPNLTLEGNVTFGEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKS 1077
QY 1080 TVVOLLERYDPLAGSVLIDGKEIKHLNVQWLRHAHIGIVSQPIILFDCSIAENIAYGNS 1139
Db 1078 TVVOLLERYDPLAGSVLIDGKEIKHLNVQWLRHAHIGIVSQPIILFDCSIAENIAYGNS 1137
QY 1140 RVYSHEEIVQAAKEANIHHFIETLPKYNTVRVGDGKTQLSGGQKORIAIARALVRQPHIL 1199
Db 1138 RVYSQEIIVRAKEANIHAFIESLPNKYSTKVGDKGTQLSGGQKORIAIARALVRQPHIL 1197
QY 1200 LLDATSALDTESEKVVQBALDKAREGRTCIIVAHRLSTIQNADLIVFQNGKVKHEGTH 1259
Db 1198 LLDATSALDTESEKVVQBALDKAREGRTCIIVAHRLSTIQNADLIVFQNGKVKHEGTH 1257
QY 1260 QLLAOKGIYFVSMVSVQAGAKR 1281
Db 1258 QLLAOKGIYFVSMVSVQAGTKR 1279


```

Db 718 AINGGLQAPAFIIFSKLIIGVTRIDDPETKQNSNLFSLFLALGIISFTTFELQGTf 777
Qy 780 GKAGEILTCLRMYVFRSMLRQDYSWEDDPKNTTGALTTRLANDAAQVKGAGSRLAVIT 839
Db 778 GKAGEILTCLRMYVFRSMLRQDYSWEDDPKNTTGALTTRLANDAAQVKGAGSRLAVIT 837
Qy 840 QNIANLGTGIIISLIYGNQTLTLLLAIVPIIAIAGVVMKMLSGOALKDKKELEGAGKIA 899
Db 838 QNIANLGTGIIISLIYGNQTLTLLLAIVPIIAIAGVVMKMLSGOALKDKKELEGAGKIA 897
Qy 900 TEAIENFTVVSLLPREKFEYMYAQSLQVPYRNSLRKAHIFGVFSFTQAMMYFSYACGF 959
Db 898 TEAIENFTVVSLLPREKFEYMYAQSLQVPYRNSLRKAHIFGVFSFTQAMMYFSYACGF 957
Qy 960 RFGAYLVANEMFQDVLVFSATVFGAMAVGVSSFAPDYAKAKVSAAHVIMIEKSPL 1019
Db 958 RFGAYLVAKMLMSFEDVLVFSATVFGAMAVGVSSFAPDYAKAKISAAHIIMIEKTPL 1017
Qy 1020 IDSYSPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLOGLSLEVKKGOTLALVSSGGCKS 1079
Db 1018 IDSYSSTGLMNTLEGNVTFGEVFNYPTRPDIPVLOGLSLEVKKGOTLALVSSGGCKS 1077
Qy 1080 TVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHLGIVSQEPILFDCSTAEINIAVDNS 1139
Db 1078 TVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHLGIVSQEPILFDCSTAEINIAVDNS 1137
Qy 1140 RVVSHEETVOAAKEANTHHTETLPEKYNTVRVGDGKTQLSGGQKORTAIARALVRQPHIL 1199
Db 1138 RVVSHEETVOAAKEANTHHTETLPEKYNTVRVGDGKTQLSGGQKORTAIARALVRQPHIL 1197
Qy 1200 LLDEATSALDTESEKVVQAEALDKAREGRTCVIAHRLSTIONADLIWVFONGVKVKEHGH 1259
Db 1198 LLDEATSALDTESEKVVQAEALDKAREGRTCVIAHRLSTIONADLIWVFONGVKVKEHGH 1257
Qy 1260 QLLAQKGIYFSMVSVQAGAKR 1281
Db 1258 QLLAQKGIYFSMVSVQAGTKR 1279

```

Search completed: November 6, 2002, 18:39:42
 Job time : 39.4943 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:45:15 ; Search time 3803.81 Seconds
(without alignments)
7047.385 Million cell updates/sec

Title: US-09-672-725C-27

Perfect score: 6473

Sequence: 1 MDPEGGRKGSAEKNFWKMGK.....LLAQKGIYFSMVSQAGAKR 1281

Scoring table: BLOSUM62

Xgapop	10.0	,	Xgapext	0.5
Ygapop	10.0	,	Ygapext	0.5
Fgapop	6.0	,	Fgapext	7.0
Delop	6.0	,	Delext	7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```

MODEL=frame2_n2n.model DEV=x1h
-Q/cgn2_lj/QPMT-spool/US09672725/runat_04112002_124341_17977/app_query.fasta_l.6435
-DB=genEmil QPMT-spool SUFFIX=rge MTNMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human.0.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT-pto -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USERS=US09672725 -CGCN_l_11409 -regat_04112002_124341_17977 -NCPU=6 -ICPU=3
-NO_XUPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

```
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_ipv:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	6473	100.0	99.9	4279	6	AX105082	Sequence
2	6466	99.9	99.9	4279	6	AX105080	Sequence
3	6461	99.8	99.8	4279	6	AX105078	Sequence
4	6456	99.7	99.7	4279	6	AX105057	Sequence
5	6447	99.6	99.6	4045	12	AF269224	Synthetic
6	6425.5	99.3	99.3	4317	4	AF045016	Canis fam
7	6425.5	99.3	99.3	4317	6	AX105059	Sequence
8	6403.5	98.9	98.9	3934	4	CPA419568	Canis fam
9	5866	90.6	90.6	3860	6	AX322787	Sequence
10	5863	90.6	90.6	4378	6	E023236	Multiidrug r
11	5862	90.6	90.6	4186	6	AX108654	Sequence
12	5859.5	90.5	90.5	4195	6	AX108656	Sequence
13	5857	90.5	90.5	3860	6	AX322789	Sequence
14	5857	90.5	90.5	4646	6	AX336420	Sequence
15	5857	90.5	90.5	4646	6	AX336708	Sequence
16	5857	90.5	90.5	4646	6	T49610	Sequence 2
17	5857	90.5	90.5	4646	9	HUMMDR1	Sequence 2
18	5857	90.5	90.5	4669	6	T08557	Homo sapien
19	5857	90.5	90.5	6505	6	AR028671	Sequence 3
20	5857	90.5	90.5	8630	6	AX012320	Sequence
21	5857	90.5	90.5	8630	6	AX012321	Sequence
22	5857	90.5	90.5	9318	6	AR028672	Sequence
23	5849.5	90.4	90.4	3988	6	AX024454	Sequence
24	5849.5	90.4	90.4	4192	9	AF016535	Homo sapi
25	5830	90.1	90.1	4669	6	AR091275	Sequence
26	5815.5	89.8	89.8	4264	6	AR051647	Sequence
27	5815.5	89.8	89.8	4264	6	AR051650	Sequence
28	5797	89.6	89.6	4669	6	AR055785	Sequence
29	5682.5	87.8	87.8	3858	4	ORU78609	Sequence
30	5664.5	87.5	87.5	4296	10	CRUFGPI	Sequence
31	5653.5	87.3	87.3	4304	10	CRUFGPI165	Sequence
32	5641	87.1	87.1	4788	6	AX323293	Sequence
33	5641	87.1	87.1	4924	10	MUSMDRIA	Sequence
34	5641	87.1	87.1	4927	10	AF257746	Rattus no
35	5638	87.1	87.1	4356	10	MUSMDRAA	Sequence
36	5619	86.8	86.8	4323	10	AF286167	Mouse multi
37	5521.5	85.3	85.3	3489	4	ABO29153	Felis cat
38	5392.5	83.3	83.3	3987	10	CRUFGPI185	Chinese ham
39	5301	81.9	81.9	4189	6	AX323291	Sequence
40	5301	81.9	81.9	4298	10	MUSMDR	Sequence
41	5291.5	81.7	81.7	4233	6	AR123273	Mouse multi
42	5280	81.6	81.6	4280	10	CRUFGPII	Sequence
43	5231.5	80.8	80.8	4254	10	RAYMDR	C.griseus P
44	4995	77.2	77.2	3682	10	MUSPGGY1A	Rat mdr MRN
45	4922.5	76.0	76.0	3924	6	AX024455	Mouse phosph

ALIGNMENTS

RESULT 1					
AXI05082					
LOCUS	AXI05082	4279 bp	DNA		
DEFINITION	Sequence 26 from Patent WO0123540.				
ACCESSION	AXI05082				
VERSION	AXI05082.1	GI:13921232			
KEYWORDS	.				
SOURCE	dog.				
ORGANISM	Canis familiaris				
					PAT 30-APR-2001 linear

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 4279)
Stocker,P.J., Steimer-Crespi,D.T., Crespi,C.L., Reif,T.C. and Patten,C.J.
P-glycoproteins and uses thereof
Patent: WO 0123540-A 26 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
source
1..4279
/organism="Canis familiaris"
/db_xref="taxon:9615"
17..3862
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37723.1"
/db_xref="gi:13921233"
/translation="MDPEGGRKSAENFWMKMKKKKKKKKKPPVSTFAMFRYSN
WLDRLYMLVGTMAAIHGAALPLMLLVGNMSTDSFANAGISRNKTFVFINESITNNT
QHFINLHEEMTVAYYVYGAGVLAAYIQVSWFGLAAGRQILKIRKQFFHAIMRQ
ETGFVDHVGELNTRLTDDYSKINEGIDKIGMFOSIATFFFTGTVTRGWKFLR
VILALSPVLGISAIIWAKILSSFTDKELLAYAKAGAAVEVLAIKRVIAFGGKKEL
ERYNNKEAGIGIKKAITANISIGRAFLLIYASALAFWGSILVSEYITIGQVL
TVFSLVGFASIOQSPSIEAFANAGAAYEIKIIDNKPSIDSYSKSGHKPDNIKG
NLEFNHVSYPSSRKEVKILKLNKLVQSGQTVALVNSGCGKSTTVOLMORLYDPFD
GMVCDGQDIRTINVRHLREITGVVQEPVLFAITAIENIRYGRNTVMDIEKAVKE
ANAYDFIKLKNPFTLVGERGAOLSGGOKORIAIARALVRNPKILLIDDEATSDLTE
SEAVVOVADKARKGRRTIVIAHRLSTVRNADVIAGDDGVIVEKGNHDELMKPKGIY
KLGTWQNRNEILENATGESKESDALEMSPKDSGLSLKRBSTERSIHAPOQDR
PTRDDSPETKQNSNMFSVLVLGIIISIFITFFLQGFTRKAGELIKRLKRMWFRSM
LRQDYSFDPKNTGALTTRLANDAQVKAIGSLRAVITQNTIANLGTIIISLYG
WOLTLALLAIVPIIAIAGVEMKMLSGALKKELEGAKIATEAENFRTVYNSLR
EQKPYMTAOSLVPRNLSKAHIFGVFSFISITOMMYFSVAGCFRFGAYLVANEFNN
FQDVLVSAIVRGMAVQVSSPADYAKAKVSAAHVIMIIIESPLIDSYSPHLKP
NTLEGNVFNENFNPYRPIPIVLQGLSLEVKGQTLALVSGSCGKSTVVOLERE
YDPLAGSLVDGKEIKHLNVQRLHGIQVSEPLIFDCSAENIATAGDSNRVYSHEE
IVQAAKEANIHHFETLPEKYNTRVGDGKQTOISGGQKORIAIARALVRNPKILLIDDEA
TSALDTESEKVVQEALDKAREGRTICIVIAHRLSTIQNADLIIVFQNGKVBHGTHQOL
LAQGIYFSMVSVOAGAKR"
BASE COUNT 1296 a 833 c 1009 g 1141 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6473.00 Matches: 1281
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-672-725C-27 (1-1281) x AX105082 (1-4279)

Qy 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPhetRlysMetGlyLys 20
Db 17 ATGGATCTCAAGAGGCCCTAAGGGAGTGCAGAGAAGAACTTCGGAAAATGGCAAA 76
Qy 21 LysSerLysLysLysGluLysLysLysLysProThrValSerThrPheAlaMetPhe 40
Db 77 AAAAGTAAAAAAGAGAGAGAAAGAAAGAACCAACTGTCCAGCAGCTTTCGAATGTTT 136
Qy 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 60
Db 137 CGCTATTCAAAATGGCTTGTATGTTGTTGGTGGGCAATGGCTGCCATCATC 196
Qy 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspPheAla 80
Db 197 CATGGAGCTGCATCCCTCTCATGATGCTGGTTTTTGGAAAACATGACATAGCTTTGCA 256
Qy 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 257 AATGCAGGAATTCAAGAAACAAACATTTTCCAGTTATAATTAATGAAAGTATTACGAAC 316

Qy 101 AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 120
Db 317 AATACACAAATTCATCAACCATCTGGAGAGGAAATACCACGATGCTATGCTATTTATTC 376
Qy 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
Db 377 AGTGGGATCGGTGCTGGCGTGGTGGTGGCTGTACATCCAGGTTTCATCTCGTGGCTG 436
Qy 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
Db 437 GCAGCAGAAAGACAGATACTCAAAATTTAGAAAACAAATTTTTCATGCTATCATGCGAC 496
Qy 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
Db 497 GAGATTGGCTGGTTGACGTGCATGCTGGGAGCTTAACACCCGGCTCACAGACGAT 556
Qy 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
Db 557 GTCTCCAAAATCAATGAAGGAATGGCCACAAAATTTGGAATGTTCTTTCATCAATAGCA 616
Qy 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
Db 617 ACATTTTCCACCGGTTTTATAGTGGGCTTTACACGCTGTTGGAAGCTTAACCCTTGTGATT 676
Qy 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
Db 677 TTGGCCATCATGCCCTGTCTTGGACTTTCACGCCCATCTGGGCAAGATACATCTTCA 736
Qy 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
Db 737 TTTACTGATAAAGAACTCTTGGCCATGCAAAAGCTGGAGCAGTAGCTGAAGAAGCTTA 796
Qy 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
Db 797 GCAGCAATCAGACTGTGATTCCTTTGGAGGACAAAGAAAGAACTTGAAGAGTACAAC 856
Qy 281 LysAsnLeuGluGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIleSer 300
Db 857 AAAATTTTAGAAGAACTAAAGAAATTTGGGATAAAGAAAGCTATCACGGCCCAACATTTCT 916
Qy 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrGlyThr 320
Db 917 ATTGTCGCGCTTTCTTATTGATCATATGCTCTGCTGCTTTCTGATGAGACC 976
Qy 321 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 340
Db 977 TCCTTGGTCTCTCCAGTGAATATACTATTGGACAAGTACTACTGCTCTCTTCTTCTGTA 1036
Qy 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGlyAlaPheAlaAsnAla 360
Db 1037 TTAATTGGGGCTTTTAGTATTGGACAGCATCCCAAGCATTTGAAGCATTTGCCAAACGCA 1096
Qy 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
Db 1097 AGAGGAGCAGCTTATGAAATCTTCAAGATAATTCACAATAAATCAACCAAGCATTTGACAGCTAT 1156
Qy 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1157 TCGAAGAGTGACATAAACACAGATAATTAAGGAAATTTGGAATTCAAAATATGTTTCAAC 1216
Qy 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1217 TTCAGTTACCTTCTCGAAAAGAGAGTTAAGATCTTAAGGGCTCTCAACCTCAAGGTTTCAAG 1276
Qy 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1277 AGTGGCAGACAGTGGCGCTGTTGGGAACAGTGGCTGCGGGAAAGAGCACACCGTGCAG 1336
Qy 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db 1337 CTGATGACAGAGGCTCTATGACCCACAGATGGCATGGTCTGTTATTTGATGGACAGACATT 1396
Qy 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnProVal 480

QY 321 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPheSerVal 340
Db 977 TCCTTGGGCCCTCCAGTGAATATCTATGGACAAGTACTACTGTCTTCTTTCTGTA 1036
QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db 1037 TTAATTGGGGCTTTAGTATTGGACAGCATCCCAAGCAATTGAAGCAATTGCAACGCA 1096
QY 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
Db 1097 AGAGAGAGAGCTTATGAATCTTCAAGATAATTGACAAATAAACCAAGCAATTGACAGTAT 1156
QY 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1157 TCGAAGAGTGACATAAACAGATATATTAAGGAAATTTGGAATTCAAAATATGTTAC 1216
QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1217 TTCAGTTACCCTTCTCGAAAAGAGTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTACAG 1276
QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1277 AGTGGCCAGACAGTGGCCCTGGTTGGGAACAGTGGCTCGCGGAAGAGCACCGCTGCAG 1336
QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db 1337 CTGATGCAGAGCTATGACCCACAGATGGCATGGTGTCTATTTGATGGACAGGACATT 1396
QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db 1397 AGGACCAATAATGTAAGGCATCTTCGGGAATTTACTGGTGTGGTGCAGTACAGGACCTGTG 1456
QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1457 TTGTTTGGCCACACGATAGCTGAAACATTCGCTATGCGCGGAAATGTCCACATGGAT 1516
QY 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1517 GAGATTGAGAAGCTGTTAAGAGACCCATGCTATGATTTATCATGAAACTACCTAAT 1576
QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
Db 1577 AAATTGACACTCTGGTTGGAGAGAGAGGGGCCAGCTGAGTGGTGACAGAAACAGAGA 1636
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuLeuAspGluAlaThr 560
Db 1637 ATCGCCATGTCTGGGGCCCTGGTTGCGAACCCCAAGATTTCTCTGCTGGATGAGGCAACG 1696
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1697 TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCTGGATGAGGCCAGAAAA 1756
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1757 GCGCGGACTACCATTTGATAGTACATCGTTTGTCTACAGTTCTGTAATGCCGATGTTCATT 1816
QY 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1817 GCTGGTTTGTGATGATGAGTCAATTTGTGGAGAAAGAAATCATGATGAACCTCATGAAAGAG 1876
QY 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1877 AANGGCATTTACTTCAAACTTGTCACAAATGCAGACAGAGGAAATGAATTTGAGTTAGAA 1936
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1937 AATGCCACTGTGTAATCCAAAAGTGAAGTGATGCCCTTGGAAATGTCTCCAAAAGATTCA 1996
QY 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db 1997 GGGTCCAGTTAATAAAGAAAGATCAACTCGCAGGAGTATACATGCACCAACCAAGGCCAA 2056

QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPheTyr 700
Db 2057 GACAGAAGCTTGGTACAAAGAGGACTTGAATGAGAATGTACTCCAGTTCTCTCTGG 2116
QY 701 ArgIleLeuLysLeuAsnSerThrGluTyrPheValValGlyIlePheCysAla 720
Db 2117 AGGATTCGAAGCTGAACCTCAACTGAATGCGCTTATTTTGTGCTGGTATATTTGTGCT 2176
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2177 ATTATAACGGAGCGCTGCAACAGCATTTTCAATAATATTTTCAAGGATATATAGGATC 2236
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2237 TTTACCCGAGATGAGGATCTCTGAACAAACAGACAGAAATAGTAACATGTTTCTGTATTG 2296
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2297 TTTCTAGTCTCTGGAATATTTCTTTTATTTACATTTTCTCTCCAGGCTTCACATTTGGC 2356
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAAGCTGGGAGATCTCTCACTAAGCGGCTTCGATACATGGTTTTCAGATCCCATGCTGAGA 2416
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGATGTGAGCTGGTGTGATGACCTTAAACACACACCTGGAGCATTGACACACAGCTT 2476
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2477 GCCAATGATCGGCTCAAGTTAAAGGGCTATAGTTCAGGCTTCTGTCAATTACCCAG 2536
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
Db 2537 AATATAGCAATCTGGGACAGCATTTATATATCTTATCTTATCTGTTGGCAATTAACA 2596
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTTACTCTTACCAATGTACCCATCATTTGCAATAGCAGGAGTGTGTTGAATGAAATG 2656
QY 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluAlaGlyLysIleAlaThr 900
Db 2657 TTGCTCGACACAGCTGAAAGATAAGAAAGAGCTAGAAGAGCTGGGAAGATTGTCTACA 2716
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2717 GAAGCCATCGAAACTTCCGAACTGTGTCTTCTTCTGCTGGGAGCAGAGTTTGAATAC 2776
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2777 ATGATGCACAGAGTTTGCAGTTACCATACAGAAACTCTTTGNGGAAACACACATCTTC 2836
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2837 GGGGTCTCATTTCTATCCAGGCAATGATGATATTTTCTTATCTGCTGCTGCTTCCGG 2896
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 980
Db 2897 TTTGGTGCCTTACTTGTGGCAATGAGTTTCATGAACCTTTCAGGATGTTCTTGTGTTATTC 2956
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
Db 2957 TCAGCTATTGTCTTTGGTCCATGGCAGTGGGCGAGTCACTGTTCACTTCTCTGACTAT 3016
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
Db 3017 GCCAAAGCCAAAGATATCAGCAGCCACGTCATCATGATCATTTGAAAAAGCCCTCTGATT 3076
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3077 GACAGCTACAGCCCTCACGGCTTCAAGCCAAATACGTTGGAAGGAAATGTGACATTTAAT 3136
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060

Db 497 GAGATTGGCTGTTGACGTGATGACGTTGGGAGCTTAACACCCGGCTCACAGACGAT 556
QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
Db 557 GTCGCCAAATCAATGAAGGAATGGCGACAAATGGAATGTTCTTCAATCAATAGCA 616
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
Db 617 ACATTTTACCAGCTTTATAGTGGGTTTACAGCTGGTGGGAAGCTAACCCCTTGAT 676
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
Db 677 TTGGCCATCAGCCCTGTTCTTGACTTTCAGCCGCCATCTGGGCAAGATACTATCTCA 736
QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu 260
Db 737 TTACTGATTAAGAACCTCTTGGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGCTT 796
QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyrAsn 280
Db 797 GCAGCAATCAGAACTGTGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAAC 856
QY 281 LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
Db 857 AAAAAATTTAGAAGAAGCTAAAGGAATTTGGGATAAAGAAAGCTATCACGGCCAACTTCT 916
QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
Db 917 ATTGGTCCCGCTTCTTATTGATCTATGCATCATATGCTCTGGCTTCTGGTATGGGACC 976
QY 321 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 340
Db 977 TCCTTGTGCTCTCCAGTGAATATTCTATTGGACAAGTACTACTGCTCTCTTTCTGTA 1036
QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db 1037 TTAATTGGGGCTTTTATGATTTGGACAGGCATCCCAAGCATTGAAGCATTTCACACGCA 1096
QY 361 ArgGlyAlaAlaTyrGluIlePheLysIleAlaAspAsnLysProSerIleAspSerTyr 380
Db 1097 AGAGGACAGCTTATGAAATCTTCAGATTAATGACAATAACCAAGCATTTGACAGCTAT 1156
QY 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1157 TCGAAGAGTGGACATAAACAGATAATATTAAAGGAAATTTGGAATTCAAAATGTTTAC 1216
QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1217 TTCAGTTACCCCTTCTCGAAAAGAAAGTAAAGATCTTAAAGGGTCTCAACCTGAAGGTTAC 1276
QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrValGln 440
Db 1277 AGTGGGCACACAGTGGCGCTGTTGGACACAGTGGCTGGGGAAGACAGCAGCCGTGCAG 1336
QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db 1337 CTGATGCAGAGGCTCTATGACCCACAGATGGCATGCTGTGATTGATGGACAGGACATT 1396
QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnProVal 480
Db 1397 AGGACCATAAATGTAAGGCATCTTCGGGAAATTACTGGTGGTGGTGGAGGCTGTG 1456
QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1457 TTGTTTGGCCACACGATGCTGAAACATTCCTATGGCCGCGAAATGTCACCATGGAT 1516
QY 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1517 GAGATTTGAGAAGCTGTTAAGGAAGCAATGCTATGATTTTATCATGAAACTACCTAAT 1576
QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlnLysGlnArg 540

Db 1577 AAATTTTGACACTCTGCTTGGAGAGAGAGGGCCAGCTGAGTGGTGACAGAAACAGACA 1636
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
Db 1637 ATCGCATTCCTCGGGCCCTGGTTGCAACCCCAAGATTCTTCTGCTGATGAGCAAG 1696
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1697 TCAGCTCTCGACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCTGGATTAAGGCCAGAAA 1756
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1757 GGCGGACTACCATTTGTGATAGCTCATCTTGTCTACAGTTCTGTAATGCCGATGTCATT 1816
QY 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1817 GCTGSGTTTGTGATGAGCTCATCTTGGAGAAAGAAATCATGATCACTCATGAAGAG 1876
QY 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1877 AAGGCGATTTACTTCAAACTTGTCAATGCAAGCAAGAGAAATGAAATTCAGTTAGAA 1936
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1937 AATGCCACTGGTGAATCCAAAAGTGAAGTGAATGCTTGGAAATGCTCCAAAAGATTCA 1996
QY 661 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 680
Db 1997 GGGTCAGCTTAAATAAAGAAAGATCAACTCGCAGAGATATACATGSCACCACAGGCCAA 2056
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
Db 2057 GACAGAAAGCTTGGTCAAAAGAGGACTTGAATGAGAAATGATACCTCCAGTTCTCTCG 2116
QY 701 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db 2117 AGGATTTCAAGCTGAAGTCAACTGAATGGCCTTATTTGGTGGTATATTTTGTGCT 2176
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2177 ATTATAACGGAGGCTTGCACACGACATTTTCAATAATATTTTCAAGGATTATAGGGATC 2236
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2237 TTTACCCGAGATGAGATCCTGAAACAAACACAGACAATAGTACATGTTTCTGTATG 2296
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2297 TTTCTAGTCTTGGAAATATTTCTTTTATACATTTTCTCCAGGGCTTCACATTTGGC 2356
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2357 AAAGCTGGGAGATCCTCATAAGCGCTTCGATACATGTTTTCAGATCCATGCTGAGA 2416
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2417 CAGGATGTCAGCTGGTTGATGACCCCTAAACACCACTGGAGCATTCACACACAGGCT 2476
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2477 GCCAATGATGGGGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGTGTCATTACCCAG 2536
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
Db 2537 AATATAGCAATCTTGGACAGGATTAATATATCCCTTAATCTATGTTGGCAATTAACA 2596
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2597 CTTTACTCTTAGCAATTTGATACCATCATTCCTCAATAGCAGAGGATGTTGTAATGAAATG 2656
QY 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluAlaGlyLysIleAlaThr 900
Db 2657 TTGCTGGACAAGCACTGAAAGATAAAGAGAGCTAGAAAGAGCTGGAGAGATTGCTACA 2716

Qy	901	GluAlaIleGluAsn	PheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr	920
Db	2717	GAAGCATCGAAACTTCGAACTGTTGTTCTTGACTCGGAGCAGAAGTTTGAATAC	2776	
Qy	921	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	940	
Db	2777	ATGTATGCACAGATTGCAAGTACCATACAGAAACTCTTTGAGGAAGACACACATCTTC	2836	
Qy	941	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	960	
Db	2837	GGGCTCTCATTTTCATACCCAGGCAATGATGATTTTCTATGCTGCTGCTTTCCGG	2896	
Qy	961	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe	980	
Db	2897	TTTGGTGCTACTTGGTGCAANTGAGTTCATGAACCTTCAGGAATGTCCTTTGGTATTTC	2956	
Qy	981	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	1000	
Db	2957	TCAGCTATTGCTTTTGGTSCCATGTCAGTGGGCGAGGTCAAGTTTCATTGCTCCTGACTAT	3016	
Qy	1001	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1020	
Db	3017	GCCAAAGCCAAAGTATCACAGCCACCGCTCATGATCATTTGATAAAAGCCCTCTGATT	3076	
Qy	1021	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1040	
Db	3077	GACAGCTACAGCCCTCACCGCCTCAAGCCAAATAGTTGGAAGAAATGTGACATTTAAT	3136	
Qy	1041	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1060	
Db	3137	GAGTGCTGTTCAACTATCCCATCGACACACATCCCGCTGCTCCAGGGCTGAGCCTC	3196	
Qy	1061	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1080	
Db	3197	GAGGTGAAGAAGGGCCAGACGCTGGCCCTCTGTAGTAGACAGTGCTGTGGAGAGCCACA	3256	
Qy	1081	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1100	
Db	3257	GTTGTTCAAGCTCTAGAGCGCTTCTATGACCCCTTGGCTGTTTCAAGTGTAAATTGATGC	3316	
Qy	1101	LysGluIleLysHisLeuAsnValGlnThrPheArgAlaHisLeuGlyIleValSerGln	1120	
Db	3317	AAAGAGATAAGACCTGTAATGTCCAGTGGCTCCGAGACACCTCGGCATCGTGTCTCAG	3376	
Qy	1121	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1140	
Db	3377	GAGCCCATCTGTTTGACTGCAGCATTCGCGAGACATTTGCTATGGAGACACAGCCGG	3436	
Qy	1141	ValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPheIle	1160	
Db	3437	GTCGTATCATGAAGAGATTATGAGGCGAGCCAAAGGAGGCAACATACACCACTTCATC	3496	
Qy	1161	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1180	
Db	3497	GAGACATCCCTGAGAAATACAAACACGAGTAGGAGACAAAGGAACCCAGCTCTCTGGT	3556	
Qy	1181	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1200	
Db	3557	GGCCAGAAACAGCGCATGGCATAGCTCGCGCTCTTGTTAGACAGCCTCATATTTTGGCT	3616	
Qy	1201	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1220	
Db	3617	TTGGATGAAGCTACATCAGCTCTCGGATACAGAAAGTGAAGAGTTGTCCAAGAGCCCTG	3676	
Qy	1221	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1240	
Db	3677	GACAAAGCCAGAGGCCGCACCTGCATTTGATGCGCCACCGCTTGTCCACCATCCAG	3736	
Qy	1241	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1260	
Db	3737	AATCGCATTTAATAGTGGTGTTCAGATGGCAAGTGCAGAGCATTTGGCACACATCAA	3796	

QY	1261	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys	1280
Db	3797	CAGCTGCTGGCCAGAAAGGCATCATATTTTCCATGTCAGTGTCCAGGCTGGAGCAAAG	3856
QY	1281	Arg	1281
Db	3857	CGC	3859
RESULT	4		
AXI05057			
LOCUS	AXI05057	4279 bp	DNA linear PAT 30-APR-2001
DEFINITION	Sequence 1 from Patent WO0123540.		
ACCESSION	AXI05057		
VERSION	AXI05057.1	GI:13921209	
KEYWORDS			
SOURCE	dog.		
ORGANISM	Canis familiaris		
REFERENCE	Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
AUTHORS	1 (bases 1 to 4279)		
TITLE	Stocker,P.J., Steinel-Crespi,D.T., Crespi,C.L., Reif,T.C. and Patten,C.J.		
JOURNAL	P-glycoproteins and uses thereof		
GENTEST	Patent: WO 0123540-A 1 05-APR-2001;		
FEATURES	Location/Qualifiers		
source	1..4279		
	/organism="Canis familiaris"		
	/db_xref="taxon:9615"		
CDS	17..3862		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC37718.1"		
	/db_xref="GI:13921210"		
	/translation="MDPEGKGSAAEKFNKMGKKKKKKKKKKKPTVTFAMFRYSN		
	WLDRLYLMLVGTMAALIHGAALPLMLVFGNMTDSFANAGISRNKTPVPIINESITNNT		
	OHFNHLEEMTYYAYYSIGIGAGVLVAAYIOVSFWCLAAAGROILKIRKQFFHAIIMRO		
	EIGWFDVHDVGLNTRTDDVDSKINEGIGDKIGNFHSHSIATFTFTGTVFGTCWKLT		
	VILASIPVLGSAIAVANTIGSSFTDKELLAYAKAGAAVEALAAIRTVIAFGGQKKEL		
	ERYKNLFEAKGIGIKKAITANTISGAAFLLIYASALAFWIGTSLVLSSEYSGVQL		
	TVFSVLVIGAFSGIQASPSIEAFANARGAAEYFKIIDNKPISDSYSKSHKHPDNING		
	NLEFKNVHFSYPSRKEVILKLNKLVQSGQTVALVNGSGCKSTTVQLMORLYDPT		
	GMVCIQGDQIRTIIVNRHLREITGVVSGQPVLFATTIAENIRYGRNVTMDIEIKAYE		
	ANVDEIMKLPNKFTLVGERGAOLSGGOKORIAARALVRNPKILLDEATKALDTE		
	SEAVQVALDKARKRRTTIVTAHSLTRVNADVIAGDDGVIVKEKNGHDEMKEGIV		
	FKLVMTQDNRNEIELEENATGKSESDALEMSPKDSGSSLLIKRSTPRSHIAPQODR		
	KLGTGTDNENVPVPSWRILKLNSTEPYFVPGFICAILINGLQLPAFSIFSRIGLI		
	FTRDPEDDKQNTNMFVSLFVLGIIISFTIFFLQGTFGKAGLEILTKRLRYWFRSM		
	LRQVDSFPDDPKNTTGALTTRIANDAAQVKGASRLVATONTIANIGLTIISLIVR		
	WQLTLLIIVPIITAIAGVBMKMSGALKDKKEGACKIATEATENFTVSVLFR		
	EQKEFMYAYQSLQVYRNSLRKAIHFGVSFTIOAMMYFVYAGCFRFGAYLVANEEFN		
	FQDVLIVPSAIVFGAMAGOVSSFPADYAKYASAAHVIMIEKSPILDSYVSHPLK		
	NLTGENTVFNENVPTRPDIPLVQLGISLEVKKGOTLALVVGSGCGCKSNVQVLLERF		
	YDPLAGSVLIDGKEIKHLNQLWRAHKGITVSOEPIFLDSCIAENIAYGDNRVYVSEH		
	IMQAKEANIHHFTETPEKYNTRVGDGTQSGGOKORIAARALVRQPHILLDLDA		
	TSALDTESEYVQEAQLEKAREGRTCVIAHRLSTIQADLLIVVFQNGKVKHEGTHQOL		
	LAQKGIYFSMVSQAQAKR"		
BASE COUNT	1294 a	834 c	1143 t

Db 2237 TTTACCGGAGATGAGGATCCTGAAACAAAACGACAGATAGTAACATGTTTCTGTATG 2296
 Qy 761 PheLeuValLeuGlyIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
 Db 2297 TTTCTAGTCCTTGGAAATATTTCTTTATATACATTTTCTCCAGGGCTTCACATTTGGC 2356
 Qy 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
 Db 2357 AAGCTGGGAGATCCTCACTAAGCGGCTTCGATACATGGTTTTCAGATCCATGCTGAGA 2416
 Qy 801 GlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
 Db 2417 CAGGATGTACGTGGTTGATGACCCCTAAACACACACCTGGAGCATTTGACAAACGAGCTT 2476
 Qy 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
 Db 2477 GCCAATGATCGGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCTGCTCATACCCAG 2536
 Qy 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleIleIleIleIleIle 860
 Db 2537 AATATAGCAATCTTGGACAGGCATTAATATATCCTTAATCTATGTTGGCAATTAACA 2596
 Qy 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
 Db 2597 CTTTACTCTTAGCAATTTGTACCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAAATG 2656
 Qy 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
 Db 2657 TTGCTGGACAGCACTGAAGATAGAAAGAGCTAGAGGAGCTGGGAAGATGCTACAC 2716
 Qy 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
 Db 2717 GAAGCCATCGAAACATTCGAACTGTTGTTCTTTCGACTCGGGAGCAGAAAGTTTGAATAC 2776
 Qy 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
 Db 2777 ATGTATGCACAGAGTTTGCAGATACCATACAGAAACTCTTTGAGGAAAGCACACATCTTC 2836
 Qy 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
 Db 2837 GGGGTCTCATTTTCTATCACCCAGGCAAGATGATATTTTCTATGCTGGCTGTTCCGG 2896
 Qy 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 980
 Db 2897 TTTGGTGCCTACTGTGTGGCAATGAGTTTCATGAACCTTTTCAGGATGTTCTTTTGGTATTC 2956
 Qy 981 SerAlaIleValPheGlyAlaMetAlaValAlaGlyClnValSerSerPheAlaProAspTyr 1000
 Db 2957 TCAGCTATTGCTTTGTGTCATGGCAGTGGGCGAGTCACTTCATTTGCTCCTGACTAT 3016
 Qy 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
 Db 3017 GCCAAAGCCAAAGTATCAGCAGCCAGCTCATCATGATCATTTCAAAAAGCCCTCTGATT 3076
 Qy 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
 Db 3077 GACAGCTACAGCCCTCAGCGGCTCAAGCCAAATACGTTGGAAGAAATGTGACATTTAAT 3136
 Qy 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
 Db 3137 GAGTGTGTGTTCAACTATCCACTCGACACAGATCCCGGTGCTCCAGGGGCTGAGCCCTC 3196
 Qy 1061 GluValLysLysGlyGlnThrLeuAlaLeuValIcylSerSerGlyCysGlyLysSerThr 1080
 Db 3197 GAGTGAAGAGGCGCAGAGCTGGCCCTCGTAGTAGCAGTGGCTGCTGGAGAGACACA 3256
 Qy 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
 Db 3257 GTTCTTCAGCTCCCTAGAGCGCTTCTATGACCCCTTGGCTGGTCACTGCTTAATGATGGC 3316
 Qy 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120

Db 3317 AAAGAGATAAAGACACCTGAATGTCCAGTGGCTCGAGCACACCTGGGCATCGTCTCTCAG 3376
 Qy 1121 GluProIleLeuPheAspCysSerIleAlaIleAlaAsnIleAlaIleValGlyAspAsnSerArg 1140
 Db 3377 GAGCCCATCCTGTTGACTGCGCAGCATTCGCGAGAACATTCGCTATGGAGACACAGCCGG 3436
 Qy 1141 ValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1160
 Db 3437 CTCGTATCACATGAAGAGATTATGAGCGAGCCCAAGGAGGCCAACATACACCATTCATC 3496
 Qy 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
 Db 3497 GAGACACTCCCTGAGAAATACACACAGAGTAGGACACAAAGAACCCAGCTCTCTGGT 3556
 Qy 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
 Db 3557 GSCCAGAAACAGCGCATTTGCCATAGCTCGCGCTCTTGTGTAGACAGCCTCATATTTGCTT 3616
 Qy 1201 LeuaspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
 Db 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAAAGCCCTG 3676
 Qy 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
 Db 3677 GACAAGCCAGAGAGCGCCACCTGCATTTGATCGCCACCGCTTGTCCACCATCCAG 3736
 Qy 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
 Db 3737 AATGAGATTTAATAGTGTGTTTTCAGAAATGCAAGAGTCAAGAGGAGCATGGCACACATCAA 3796
 Qy 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
 Db 3797 CAGCTGCTGGCCGAGAAAGGCATCTATTTTCCATGTTGTCAGCTGGAGGAGGAGAAAG 3856
 Qy 1281 Arg 1281
 Db 3857 CGC 3859
 RESULT 5
 AF269224 4045 bp DNA linear SYN 04-JUL-2000
 LOCUS Synthetic construct Canis familiaris his-tagged-multidrug
 DEFINITION resistance glycoprotein gene, complete cds.
 ACCESSION AF269224
 VERSION AF269224.1 GI:8926216
 KEYWORDS
 SOURCE
 ORGANISM
 synthetic construct.
 synthetic construct
 artificial sequence.
 REFERENCE
 1 (bases 1 to 4045)
 AUTHORS Roulet,A.C., Puel,O., Gesta,S., Alvinerie,M. and Pineau,T.J.
 TITLE Analysis of dog MDRI P-glycoprotein
 JOURNAL Unpublished
 REFERENCE
 2 (bases 1 to 4045)
 AUTHORS Roulet,A.C., Puel,O., Gesta,S., Alvinerie,M. and Pineau,T.J.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAY-2000) Pharmacology, INRA, 180 Chemin de
 Tournefeuille, BP3, Toulouse 31931, France
 FEATURES
 location/Qualifiers
 1..4045
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="original mRNA isolated from Beagle dog colon"
 101..3964
 /codon_start=1
 /transl_table=11
 /product="his-tagged-multidrug resistance glycoprotein
 MDRI"
 /protein_id="AAF81747.1"
 /db_xref="GI:8926217"
 /translation="MHHHHHDPGGGRKGSAEKNFMKMGKSKNEKKKPTVSTFA
 MFRYSNLDRLYMLVGTGMAALIHGAALPLMMLVFGNMTDFPANAGISRNKTFPVIIINE
 SITNNTQHFINHLEEMTYYAYYSIGAGVLVAAYIQVSWFCLAGRLILKIRKQFF

HAIMRQIGWFDVHDVGLNLRITDDVSKINEGIDKIMGFOSIAFTFFFTGVGTFTR
 GWKLTVLIAISPLVGLSAAIGKAIWKLISLFTDKELLAAYAKAGAAEEVLAIRVIAFG
 GOKKELERYKNLEAKRAGIKKAITANISIGRAFLILLIYASIALAFWGSILVLSSEY
 SIGOVLVFVSVLIGAFSIGQASPSIEAFANARGAAYEIFKIIDNKPSIDSYSKSGHK
 PDNKGMEKRNHFSPSKREVKILKGLNKVOSQOTVALVNSGCGSKSTTVOLMQR
 LYDPTGMVCIQDQDIRTINVRHLREITGVSOEPLVFATTAENIRYGRNVTMDER
 EKAVKEANAYDFIMKLPNKFDTLVGERGAOLSGQKORIAIALVRNPKILLLDEAT
 SALDTESEVVOVALDKARKGRRTTIVIAHRLSTVRNADVTAGDDGVIVPKGNHDELM
 KEKIGYKLVMTQTRNEIELENAIGESKESDALPMSPKDSGLSKRSTRSRTHA
 PQGDRLKGYKEDLNENVPVFWNRILKLNSTWPIVFGICAILNGGLQPAFSIIF
 SRLIGIFTRDEPETKRONMFSVLFLVIGIISFTFFLQGFQKAGBILTKRLRY
 MYFMSRUDQWFPDDPNTGALTTRLANDAQAQVKAIGSRLAVTQNTIANIGTGI
 ISLYIGWQLTLLALLIPIIAGVEMKMLSGQALDKKELEGACKIATEAENFRT
 VVSLTRFQEDVLLVFAISVFCAMAGOVSSPAPDYAKAKYSAAHVIMLBKSPLLDSY
 ANEFMPODVLVFAISVFCAMAGOVSSPAPDYAKAKYSAAHVIMLBKSPLLDSY
 PHGLKPNLTSGNTEVENVPYTRPDIPVLOGLSLEVKKGOTLALVSGSGKSTVW
 OLLERFYDPLAGSVLIDCKEIKHLNQLWLRHIGVISOEPLFDCSIAENIAYGDNRS
 VVSHREIMAAKCANIHFFIETLPERKINTRVGDGKQLSGGQKORIAIALVRPHI
 LLDDEATSAIDTESKRVQOALDKAREGRTCTIVIAHRLSTIQNALIVVVFQNGKVEH
 GTHQQLIAQKGIYFSMISVQAGAKR*

BASE COUNT 1171 a 850 c 986 g 1038 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0 Length: 4045
 Score: 6447.00 Matches: 1275
 Percent Similarity: 99.84%
 Best Local Similarity: 99.61% Conservative: 3
 Query Match: 99.60% Mismatches: 2
 DB: 12 Indels: 0 Gaps: 0

US-09-672-725c-27 (1-1281) x AF269224 (1-4045)

QY 2 AspProGluGlyArgLysGlySerAlaGluLysAsnPhetPrLysMetGlyLysLys 21
 Db 122 GATCCTGAAGAGGCCCTAAGGGAGTGCAGAGAAGAACTTCTGGAAAATGGGCAAAAA 181
 QY 22 SerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPheArg 41
 Db 182 AGTAAAAAATGAGAGAAGAAAGAAACCAACCTCTCAGACGCTTGGCAATGTTTCCG 241
 QY 42 TyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLalleleHis 61
 Db 242 TATTCAAATTTGGCTGATAGTGTGTATATGTTGGGGACAATGCCATGCTCATCCAT 301
 QY 62 GlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAsn 81
 Db 302 GGAGCTGCACCTCCTCATGATGCTGGTGTGGTGGGACAATGACATAGCTTTGCCAAT 361
 QY 82 AlaGlyIleSerArgAsnLysThrPheProValIleAlaAsnGluSerIleThrAsnAsn 101
 Db 362 GCAGGAATTTCAAGAAACAAAACCTTTTCCAGTTATATTAATGAAAGTATTACGAACAT 421
 QY 102 ThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrSer 121
 Db 422 ACACACATTTTCATCAACCATCTCGAGGAGGAAATGACCACGCTATCCCTATTATTACAGT 481
 QY 122 GlyIleGlyAlaGlyValLeuValAlaLalaTyrIleGlnValSerPheTrpCysLeuAla 141
 Db 482 GGGATCGGTGCTGGCTGCTGGTGGCTGTATCATCCAGGTTTCATCTGGTGGCTGGCA 541
 QY 142 AlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGlu 161
 Db 542 GCAGGAAGACAGATCTCAAAATAGAAAAAATTTTTCATGCTATCATGCGACAGGAG 601
 QY 162 IleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspVal 181
 Db 602 ATTGGCTGGTTGACGTGATGACGTGGGGAGCTTAACACCCGGCTCACAGCATGTC 661
 QY 182 SerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAlaThr 201
 Db 662 TCCAAATCAATGAAGGAATTTGGGCNCAAAATTTGGAATGTTCTTTTCAATCAATGACACA 721

QY 202 PhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIleLeu 221
 Db 722 TTTTTCACCGGTTTATAGTGGGGTTTACACGTGGTGGGAAGCTAACCCCTTGTGATTTCG 781
 QY 222 AlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPhe 241
 Db 782 GCCATCAGCCCTGTTCTTGACCTTTAGCCGCCCATCTGGCGAAGATATCTATCTCATTT 841
 QY 242 ThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeuAla 261
 Db 842 ACTGATAAAGAACTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTAGCA 901
 QY 262 AlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsnLys 281
 Db 902 GCAATCAGAACTGTGATGGCTTTGGAGGACAAAAGAAAGAACTTGAAAGGTACAAACAA 961
 QY 282 AsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSerIle 301
 Db 962 AATTTAGAAGAACTAAAGAAATGGGATAGAAAGAACTATCAGGCCCAACATTTCTATT 1021
 QY 302 GlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSer 321
 Db 1022 GGTCCGCTTCTTATTATGATCTATGCATCATATGCTCTGGCTTCTGGTATGGACCTCC 1081
 QY 322 LeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerValLeu 341
 Db 1082 TTGGTCTCTCCAGTGAATATTCTATTGGACAAGTACTCCTCTCTCTTTCTGTATTATA 1141
 QY 342 IleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArg 361
 Db 1142 ATTGGGCTTCTTAGTATTGGACAGCATCCCAAGCAATTTGAAGCAATTTCAAAATTT 1201
 QY 362 GlyAlaAlaTyrGluIlePheLysIleLeuAspAsnLysProSerIleAspSerTyrSer 381
 Db 1202 GGAGCAGCTTATGAATCTTCAAGATAATTGACATAAACCACCAAGCATTGACAGCTATTTCG 1261
 QY 382 LysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPhe 401
 Db 1262 AGAGTGGACATTAACCAAGATAATATTAGGGAAAATTTGGAATTTCAAAATTTTCCACTTC 1321
 QY 402 SerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLysValGlnSer 421
 Db 1322 AGTTACCTTCTCGAAAAGAAAGCTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTCCAGAGT 1381
 QY 422 GlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeu 441
 Db 1382 GGGCAGACAGTGGCGCTGGTTGGGAACAGTGGCTGCGGGAAGAGCAGCAGCGCTGAGCTG 1441
 QY 442 MetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArg 461
 Db 1442 ATGCAGAGGCTCTATGACCCCAACAGATGGCATGGTCTGTATTGATGGAGACAGATTAGG 1501
 QY 462 ThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeu 481
 Db 1502 ACCATAAATGTAAAGCATCTTCGGGAAATTAAGTGGTGGTGGTGGTGGTGGTGGTGGT 1561
 QY 482 PheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAspGlu 501
 Db 1562 TTTGGCCACAGATAGCTGAAAACATTCGCTATGGCCGCAAAATTCACCATGATGATGAG 1621
 QY 502 IleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsnLys 521
 Db 1622 ATTGAGAAAGCTGTTAAGGAAGCCCAATGCCATGATGATTTATCATGAAACTACCTTAATAA 1681
 QY 522 PheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArgIle 541
 Db 1682 TTTGACACTCTGGTGGAGAGAGAGGGGCCCGCCAGCTGAGTGGTGGTGGAGACAGAGATC 1741
 QY 542 AlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAlaThrSer 561
 Db 1742 GCCATTTGCTGGGGCCTGGTTCGCAACCCCAAGATTTCTTCTGCTGGATGAGGCAACGTCA 1801
 QY 562 AlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLysGly 581

1802 GCTCTGGACACTGAAGCTCAACAGCTGGTTACAGTGGCCCTGGATAAGGCCAGAAAGGC 1861
582 ArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla 601
1862 CGGACTACCATTGTGATAGCTCATCGTTTGTCTACAGTTCTGTAATGCCGATGCTATTGCT 1921
602 GlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGluLys 621
1922 GGTTCATGATGATGAGTCAATGTGGAGAAAGGAATCATGATGAATCATGAAGAGAGAG 1981
622 GlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGluAsn 641
1982 GGCATTTACTTCAAACTTGTCACAATGCACAGACAGAGGAATGAAATTCAGTTAGAAAT 2041
642 AlaThrGlyLysSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSerGly 561
2042 GCCACTGGTGAATCCAAAGTGAAGATGATGCTTGGAAATGTCTCCAAAGATTCAAGG 2101
562 SerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlnAsp 681
2102 TCCAGTTTAATAAAAAAGAGATCACTCGCAGAGATATACATGCACACACAGCCAGAC 2161
682 ArgLysLeuGlyThrLysGluAspLeuAsnGluValProValSerPheThrArg 701
2162 AGAAGCTTGGTACAAAGAGGACTTGAATGAGAAATGTACCTCCAGTTTCTCTGGAGG 2221
702 IleLeuLysLeuAsnSerThrGluTrpProThrPheValIleGlyIlePheCysAlaIle 721
2222 ATTCGAAGCTGAATCAACTGAATGGCCTATTATTTGTGGTGTATATTTGTGCTATT 2281
722 IleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePhe 741
2282 ATAACCGAGGCTGCAACACGATTTTCAATAATATTTCAAGGATATAGGGATCTTT 2341
742 ThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe 761
2342 ACCCGAGATGAGGATCTCTGAACAAACAGACAGAAATAGTAACATGTTTCTGTATTGTTT 2401
762 LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys 781
2402 CTAGTCTCTGGAATATTCTTTTATTACATTTTCTCCAGGCTTCACATTTGGCAAA 2461
782 AlaGlyGluIleLeuThrLysArgLeuArgLysMetValPheArgSerMetLeuArgGln 801
2462 GCTGGGAGATCCTCACTAAGCGCTTCGATACATGTTTTCAGATCCATGCTGAGACAG 2521
802 AspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla 821
2522 GATGTCAGCTGGTTTGATGACCTTAAACACCACTGGAGCATTTGACAAACCGCTTGCC 2581
822 AsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn 841
2582 AATGATCGGCTCAAGTTAAAGGGCTTATAGTTCCAGGCTGCTGTCATTAACCAAGAT 2641
842 IleAlaAsnLeuGlyThrGlyIleIleSerLeuIleThrGlyTrpGlnLeuThrLeu 861
2642 ATACCAATCTTGGACAGGCAATTTATATCTTAAATCTATGTTGGCAATTAACACTT 2701
862 LeuLeuLeuAlaIleValProIleAlaIleAlaIleAlaGlyValValGluMetLysMetLeu 881
2702 TTACTCTTAGCAATTTGACCAATCATTTGCAATAGCAGGAGTTGTTGAAATGAAATGTTG 2761
882 SerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGlu 901
2762 TCTGGACAAACCTGAAAGATPAGAAAGCTAGAGAGCTGGAAGATTGCTACAGAA 2821
902 AlaIleGluAsnPheArgThrValIleSerLeuThrArgGluGlnLysPheGluTrpMet 921
2822 GCCATCGAAAACCTCCGAACCTGTTGTTCTTGTGATCGGGAGCAGAAAGTTGATATCATG 2881
922 TyrAlaGlnSerLeuGlnValProThrArgAsnSerLeuArgLysAlaHisIlePheGly 941

RESULT 6
AF045016
LOCUS

AF045016

4317 bp

mRNA

linear

MAM 07-FEB-1998

2882 TATGCACAGAGTTTGCAGTACCACATACAGAAACTCTTTTGGAGAAAGCACACATCTTCGGG 2941
942 ValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPhe 961
2942 GTCTCATTTTCTATCACCCAGGCAATGATGATTTTTCTTCTATGCTGCTGCTGTTCCGGTTT 3001
962 GlyAlaTyrLeuValAlaAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSer 981
3002 GGTGCTACTTGTGTGGCAAAATGAGTTTCATGAACCTTTCAGGATGTTCTTTGGTATTCTCA 3061
982 AlaIleValPheGlyAlaMetAlaValGlnValSerSerPheAlaProAspTyrAla 1001
3062 GCTATTGCTTGTGGTCCATGGCAGTGGGAGGTGAGTTCATTGCTCTGCTGACTATGCC 3121
1002 LysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIleAsp 1021
3122 AAAGCCAAAGTATCAGCAGCCACGTCATCATGATCATTTGAAAAAGCCCTCTGATTGAC 3181
1022 SerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGlu 1041
3182 AGCTACAGCCCTCAGGCCCTCAAGCCAATACGTTGGAAGGAAATGTGACATTTTAATGAG 3241
1042 ValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeuGlu 1061
3242 GTCTGTGTTTCAACTATCCCACTCGACACACATCCCGTGTCTCCAGGGGCTGAGCCTCGAG 3301
1062 ValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrVal 1081
3302 GTGAAGAAGGGCCAGACGCTGCCCTCTAGGTAGCAGTGGCTGTGGGAAGACACAGTT 3361
1082 ValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGlyLys 1101
3362 GTTCAGCTCTAGACGCTTCTATGACCCCTTGGCTGGTTCAGTGCTAATTTGATGGCAAA 3421
1102 GluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGlu 1121
3422 GAGATAAAGCACCTCAATGTCAGTGGCTCCGAGCACACCTGGGCATCGTCTCTCAGGAG 3481
1122 ProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgVal 1141
3482 CCCATCTCTGTTGACTGCGACATTCGCGAGAACATTTGGCTATGGAGAACAGCCGGGTC 3541
1142 ValSerHisGluGluIleValGlnAlaLysGluAlaAsnIleHisHisPheIleGlu 1161
3542 GTATCACATGAAGAGATTATGCGAGCGCCAGGAGGCCAACATACACCACTTCATCGAG 3601
1162 ThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGly 1181
3602 ACATCCCTCGAGAAATACAAACACAGAGTAGGAGACAAAGGAACCCAGCTCTCTGGTGGC 3661
1182 GlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeu 1201
3662 CAGAAACAGCGCATTTGCCATAGCTCGCGCTCTTGTGTAGACGCCCTCATATTTGCTTTTG 3721
1202 AspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGlnAlaLeuAsp 1221
3722 GATGAAGCTACATACACTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAAGCCCTGGAC 3781
1222 LysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsn 1241
3782 AAAGCCAGAGAAAGGCGCGCACCTGTCATTTGTATCGCCACCGCTTGTCCACCATCCAGAA 3841
1242 AlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGlnGln 1261
3842 GCAGATTTAATAGTGTGTTTTCAGAAATGGCAAGTCAAGAGCATGGGCACATCAACAG 3901
1262 LeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLysArg 1281
3902 CTGCTGGCTCAGAAAGGCATCTATTTTTCATGATCAGTGTCCAGGCTGAGGCTGAGCAAAACGC 3961

DEFINITION Canis familiaris multidrug resistance p-glycoprotein (MDR1) mRNA,
complete cds.
ACCESSION AF045016
VERSION AF045016.1 GI:2852440
KEYWORDS dog.
SOURCE Canis familiaris
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 4317)
AUTHORS Puel,O., Lepage,J.F., Alvinerie,M., Galtier,P. and Pineau,P.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1998) Pharmacology, INRA, BP 3, 180 Chemin de
Tournefeulle, Toulouse Cedex 9 31931, France
FEATURES
source
1..4317
/organism="Canis familiaris"
/db_xref="taxon:9615"
/tissue_type="colon"
/note="beagle"
1..4317
/gene="MDR1"
70..3912
/gene="MDR1"
/codon_start=1
/product="multidrug resistance p-glycoprotein"
/protein_id="AAC02113.1"
/db_xref="GI:2852441"
/translation="MDEGGRKSGAEKFNWKGKSKKKEKKKPPVSTFAMPRYNW
LDRMLVGTMAALIHGAALPMLVFGNMTDSFANAGISRNKTFPVIINESITNTQ
HFINHEEEMTYAYYSIGIGAGLVAAIYQVSWFGLAAGRLKIRKQFFHAIMRQE
IGFNDHVGELATRLTDVSKINEGDKVGFQSIATFFFTGTVGTPGWKLLV
ILAIISPLVGSAAIWAIRKISFTDKELLAYAKAGAAVEVLAIRTVIAFGOKKLE
RYNKLNEAKRIGIKKAITANISIGAAFLIYASALAFWYGTSLVLSSEYITIGQVLT
VFVSLVIGARSIGQASPSIEAFNARGAAEYIFKIDNRPISDSKSGHKPDNIAGN
LEFNHFSIPSRKRLKILGNLKVQSGQTVVALVNSGCGSKSTTVOLMORLYDPTDG
MVCIDGDIITINVRHLREITGVVQSEPVLFATIAENIRYGRNVTDEIEKAVKEA
NAYDFIMKLPKRGDTLVGERGRLSGQKQRIARALVRNPKILLIDDEATSLDYES
EAVQVALDKARGRITVIAHRLSTVRNADVIAGFDGVIYVKEGNHDELMKEKGIYF
KLVTMTRGNEIELENATGSKSEDALEMSPKDSGLIKRSTERSIHAQGDQRK
LSTKEDLNVPNSVFWRLILKNSTENPFPVVGIFCAILINGLQPAFSIIFSRIGIF
TREDDETRKNQSNMFSLVFLVIGIISFTFFLOQTFGKAGEILTKRLRYWFRSML
RQVSWFDDPTKATLRLANDAAQVKAIGSLRAVITONIANIGTIIISLIYWG
QLLLLLIAPVITAIAGVVMKMLGQALKKELEGACKIATELENFTVVSLETR
QKFEYVTAQSLPYRNLSLRKAHIFGVFSITQAMVFSYAGCFRGAYLVANEFNF
QDVLVFSATVFCMAVGVSSFPADYAKAKYSAAHVIMLIEKSLPIDSYSVHGLKPN
TLEGNVTFNEVFNPTREDIPVLQGLSLEVKGQILALVSGSGCKSTVQQLERP
DLPLAGSVLDGKEIKHLNVLRAHLGIVSQEPILFDCSIAENIAYGDNRSVVSHEEI
MQAEKANIHFFIETLPEKYNTRVGDGKQQLSGGQKQRIARALVRPHILLIDDEAT
SLDPTSEKSVQBALDKAREGRTICIVIAHRLSTIQNADLIIVVFQNGKVEHGTHOOLL
AOKGIYFSMISVQAGAKR"

BASE COUNT 1293 a 844 c 1019 g 1161 t
ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	4317
Score:	6425.50	Matches:	1273
Percent Similarity:	99.69%	Conservative:	4
Best Local Similarity:	99.38%	Mismatches:	3
Query Match:	99.27%	Indels:	1
DB:	4	Gaps:	1

US-09-672-725c-27 (1-1281) x AF045016 (1-4317)

Qy	1	MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys	20
Db	70	ATGGATCCTGAAGGAGCCGTAGGGAGTGACAGAGAAAGCACTCTGGAAATGGCCAA	129
Qy	21	LysSerLysLysLysGluLysLysGluLysLysProthrValSerThrPheAlaMetPhe	40
Db	130	AAAAGT---AAAAGAGAGAGAAAGAAACCAACTCTCAGCACGTTTGGCAATGTTT	186
Qy	41	ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle	60

Db	187	CGCTATTCAATGGCTTGATAGTTGTATATGTTGGTGGGACAAATGGCTGCATCATC	246
Qy	61	HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	80
Db	247	CATGGAGCTGCACCTCCCTCATGATGCTGGTTTTGGAAACATGACAGATAGCTTTGCA	306
Qy	81	AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	100
Db	307	AATGCAGGAATTCAGAAACAAACCTTTTCCAGTTATATAATTAATGAAGTATTACGAAC	366
Qy	101	AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyrTyr	120
Db	367	AATACACAACATTTTCATCAACCATCTGCAGGAGGAATGACCACGATGCTATTATTAC	426
Qy	121	SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu	140
Db	427	AGTGGGATCGGTGCTGGCTGCTGGTGGCTGTATCATCCAGGTTTCATTTCTGGTCTG	486
Qy	141	AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln	160
Db	487	GCAGCAGGAAGACAGATACTCAAAATTAGAAACAATTTTTTCATGCTATCATGCCACAG	546
Qy	161	GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp	180
Db	547	GAGATTGGCTGGTTTGACGTGCATGACGTTGGGAGCTTAACACCGCGCTCACAGACGAT	606
Qy	181	ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla	200
Db	607	GTCTCCAAATCAATGAAGAAATGGCGACAAAGTTTGAATGTTCTTTCAATCAATAGCA	666
Qy	201	ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle	220
Db	667	ACATTTTTACCCGGTTTTATAGTGGGGTTTTACACCTGGTGGGAAGCTAACCCCTTGAT	726
Qy	221	LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer	240
Db	727	TTGGCCATCAGCCCTCTCTTGACATTTTCAGCGGCCATCTGGSCAAAGATACACTTCTCA	786
Qy	241	PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu	260
Db	787	TTTACTGATAAAGAACTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGCTTTA	846
Qy	261	AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysLysLysLysLysLys	280
Db	847	GCAGCAATCAGAACTGTGATTGCCCTTGGAGGACAAAGAAAGAACTTGAAGGTTACAC	906
Qy	281	LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer	300
Db	907	AAAAATTTAGAAAGCTAAAGAAATTTGGATAAAGAAAGCTATCACGCCCAACATTTCT	966
Qy	301	IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr	320
Db	967	ATTGGTGGCGCTTCTTATTTGATCTATGATCATATGCTCTGGCTTCTGGTATGGGACC	1026
Qy	321	SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal	340
Db	1027	TCCTTGGTCTCTCCAGTGAATATACTATTTGCACAGGTACTCACTGCTCTCTTTCTGTA	1086
Qy	341	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	360
Db	1087	TTAATTTGGGCTTTTATAGTATTGGACAGGCATCCCAAGCAATTTGAAGCAATTTGCAACGCA	1146
Qy	361	ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	380
Db	1147	AGAGAGCAGCTTATGAAATCTTCAAGATAATTTGACATAATTAACCAAGCATTCACAGCTAT	1206
Qy	381	SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	400
Db	1207	TGCAAGAGTGGACATTAACCAATATATAATTAAGGAAATTTGGAATTTCAAAATGTTTCAAC	1266
Qy	401	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	420

Db 1267 TTCAGTTACCCCTCTCGAAAAAGAGTTAAGATCTTTAAAGGGTCTCAAACTGAAGTTTCAG 1326
Qy 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1327 AGTGGGCAGACAGTGGCGTGGTTGGGAACACTGGCTGGGGAAGAGCACGACCGTGCAG 1386
Qy 441 LeuMetGlnArgLeuThrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db 1387 CTGATGCAGAGGCTCTATGACCCACAGATGGCATGGTCTGTATTGATGGACAGGACATT 1446
Qy 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db 1447 AGGACCATAAATGAAGGCATCTTCGGAAATTAAGTGGTGGTGGTGGTGGTGGTGGTGG 1506
Qy 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1507 TTTGTTGGCCACCACGATAGCTGAACACATTCGCTATGCGCGCGGAATATGTCACCATGAT 1566
Qy 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1567 GAGATTGAGAAAGCTCTTAAGGAAGCCATGCTATGATTTTATCATGAAACTACCTAAT 1626
Qy 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysArg 540
Db 1627 AAATTTGACACTCTGGTTGGAGAGAGGGGCGCGCTGAGTGGTGGACAGAAACAGAGA 1686
Qy 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
Db 1687 ATCGCCATTGCTCGGCGCCCTGGTTCCCAACCCCAAGATTCTTCTGCTGGATGAGGCAAG 1746
Qy 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1747 TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGTGGCCCTGGATAAAGGCCAGAAA 1806
Qy 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1807 GCGCGGACTACCAATTGTGATAGCTCATCGTGTGTACAGTTGCGTAATCCGATGTCATT 1866
Qy 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1867 GCTGGTTTTGATGATGGATCATTTGGAGAAAGGAAATCATGATGAATCATGAAGAG 1926
Qy 621 LysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1927 AAGGGCATTTACTTCAAACTTGTCAAAATGCACAAAGAGGAAATGAAATTGAGTTAGAA 1986
Qy 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1987 AATGCCACTGGTGAATCCAAAAGTGAAGTGTGCTTGGAAATGTCTCCAAAAGATTCA 2046
Qy 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db 2047 GGGTCCAGTTTAAATAAAGAAGATCAACTCGCAGAGATATACATGCCACCAAGSCCAA 2106
Qy 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPheThrPro 700
Db 2107 GACAGAAAGCTTGGTACAAAAGAGAGCTTGAATGAGAATGTTCTCCTTCAGTTTCTCTGG 2166
Qy 701 ArgIleLeuLysLeuAsnSerThrGluThrProTyrPheValIleGlyIlePheCysAla 720
Db 2167 AGGATTCGAAAGCTGAACTCAACTGAATGGCCCTTATTTTGTGGTGGTATATTTGTGCT 2226
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2227 ATTATAAACGGAGCCCTGCACACGACATTTTCAATAATATTTCAAGGATTTATAGGATC 2286
Qy 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2287 TTTACCGGAGATGAGGATCTCGAAACAAAACGACAGAAATAGTAACATGTTTCTGTATTGG 2346
Qy 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2347 TTTCTAGTCTTGGATTTATTTCTTTTATTATACATTTTCTCCAGGGCTTCACATTTGGC 2406

Qy 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2407 AAAGCTGGGAGATCCTCACTAAGCGCTTCGATACATGGTTTTCAGATCCATGCTGAGA 2466
Qy 801 GlnAspValSerTyrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2467 CAGGATGTCAGCTGGTTTGATGACCCCTAAAACACCACTGGAGCATTTGACACACAGCCTT 2526
Qy 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2527 GCCAATGATGGCGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTGTGTCATTACCCAG 2586
Qy 841 AsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTyrGlnLeuThr 860
Db 2587 AATATAGCAATCTTGGACAGGCATTTATATATCTTAATCTATGTTGGCAATTAAACA 2646
Qy 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2647 CTTTTACTCTTTAGCAATGTGTACCCATCATTCGAATAGCAGAGTTGTGAATGAAATG 2706
Qy 881 LeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
Db 2707 TTGTCTGGACAAGCACTGAAGATAAAGAGACTAGAAGAGCTGGGAAGATTGCTACA 2766
Qy 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2767 GAAGCCATCGAAAACTTCCGAACCTGTTGTTTCTTGTGACTCGGGAGCAGAAATTTGAATAC 2826
Qy 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2827 ATGTATGCACAGAGTTTGAAGTACCATACACAGAAACACTTTTGGAGAAAGCACATCTTC 2886
Qy 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2887 GGGTCTCATTTTCTATCACCCAGCAATGATGATATTTTCTCTATGCTGGCTGTTCCGG 2946
Qy 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 980
Db 2947 TTTGGTGCTACTTGGTGGCAAAATGAGTTTCATGAACATTCAGGATGTTCTTTTGGTATTC 3006
Qy 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
Db 3007 TCAGTATGTGCTTTGGTGCCATGGCAGTGGGAGGTCAGTTCATTTGCTCCTGACTAT 3066
Qy 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleLeuLysSerProLeuIle 1020
Db 3067 GCCAAAGCCAAAGTATCAGCAGCCACGTCATCATGATCATTTGAAAAAGCCCTCTGATT 3126
Qy 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3127 GACAGCTACAGCCCTCAGCGCTCAAGCCAAATACGTTGGAAGGAAATGTGACATTTAAT 3186
Qy 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3187 GAGTCTGTTTCAACTATCCCACTCGACACACATCCCGTGTCTCCAGGGGCTGAGCCTC 3246
Qy 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
Db 3247 GAGGTGAAGAAGGGCCAGACGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGAGACACA 3306
Qy 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
Db 3307 GTTGTGTAGCTCTTAGAGCGCTTCTATGACCCCTTGGCTGGCTCAGTGTCAATTGATGGC 3366
Qy 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120
Db 3367 AAAGAGATAAAGACCTCAATGTCCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAG 3426
Qy 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
Db 3427 GAGCCCATCTCTTTGACTGCGAGCATTCGCCAGAAACATTTGCTATGGAGACAACAGCCGG 3486

Qy	261	AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn	280
Db	847	GCAGCAATCAGAACTCTGATTGCCTTTGGAGGACAAAAGAAAGAACTTGAAGAGTCAAC	906
Qy	281	LysAsnLeuCluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer	300
Db	907	AAAAATTTAGAAGAGCTAAAAGAAATGGGATAAAGAAGGCTATCAGCGGCAACATTCT	966
Qy	301	IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrGlyThr	320
Db	967	ATTGGTCCGCTTCTATTATGATCATCATATGCTCTGGCTTCTGGTATGGAGCC	1026
Qy	321	SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal	340
Db	1027	TCCTTGGTCTCTCCAGTGAATATACTATTGGACAGGTACTCACTGCTCTCTTTCTGTA	1086
Qy	341	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGlyAlaPheAlaAsnAla	360
Db	1087	TTAATTGGGCTTTTAAATTTGGACAGGCATCCCAAGCATTTGAAGCATTTGCACAGCA	1146
Qy	361	ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	380
Db	1147	AGAGGACAGCTTATGAAATCTTCAAGATAATTTGACATAAACAACCAAGCATTTGACAGCTAT	1206
Qy	381	SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	400
Db	1207	TCGAAGAGTGGACATTAACACAGATATATTAAAGGAAATTTGGAAATTCAAAAATGTTCCAC	1266
Qy	401	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	420
Db	1267	TTCAAGTTACCCCTTCTCGAAAAGGTTAAGATCTTTAAAGGCTCTCAACCTGAAGGTTCCAG	1326
Qy	421	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln	440
Db	1327	AGTGGGACAGACTGGCGCTGGTTGGAAACAGTGCTCGGGGAAGACACACCGTCGAC	1386
Qy	441	LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	460
Db	1387	CTGATGCAGAGGCTCTATGACCCACAGATGCATGGTCTGTATTGATGGACAGGACATT	1446
Qy	461	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	480
Db	1447	AGGACCATAATCTGAAGGCATCTCGGGAATTTACTGTGTGGTGAGTCAGGAGCCGTGTG	1506
Qy	481	LeuPheAlaThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	500
Db	1507	TTGTTTGCACACAGTAGCTGAAAACATTCGCTATGCGCGCGAAAAATGTCAACCATGGAT	1566
Qy	501	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	520
Db	1567	GAGATTGAGAAAGCTGTTAAGGAAGCAATGCCTATGATTTTATCATGAACACTACCTTAAT	1626
Qy	521	LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg	540
Db	1627	AAATTTGCACACTCTGGTTGGAGAGAGAGGGCGCCGCTGAGTGGTGGACAGAAACAGAGA	1686
Qy	541	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	560
Db	1687	ATCGCCATTGCTGGGCGCTGGTTGCGAACCCCAAGATCTTCTGCTGGATGAGGCAACG	1746
Qy	561	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	580
Db	1747	TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGTGGCGCTGGATAGGCCAGAAAA	1806
Qy	581	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	600
Db	1807	GCCCGGACTACCATTTGTGATGCTCATCTGTTGCTACAGTTCGTAAATGCCGATGTCTATT	1866
Qy	601	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	620
Db	1867	GCTGGTTTGATGATGGAGTCATTGTGGAGAAGGAATATCATGATGAACACTCATGAAGAG	1926
Qy	621	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleLeuGlu	640

1927	Db		AAGGGCAATTAC	TTCAAACTTGT	CAACATG	CACAGACAGAGAAATGA	AAATGAGTTAGT	GAA	1986	
641	Qy		AsnAlaThrGly	GluSerLys	SerGluSer	AspAlaLeu	GluMetSer	ProLysAspSer	660	
1987	Db		AATGCCACTG	TGAATCCAAA	AGTGAAGCTG	ATGCTTGGAAATGTCT	CCAAAAGATTCA		2046	
661	Qy		GlySerSerLeu	leuLysArg	SerThrArg	ArgSerIle	HisAlaPro	GlnGlyGln	680	
2047	Db		GGGTCCAGTT	TAATAAAAA	AGATCAACTC	CCAGGAGTATAC	ATGACACAC	AAAGGCCAA	2106	
681	Qy		AspArgLysLeu	GlyThrLys	GluAspLeu	AsnGluAsn	ValPro	ProValSerPheTrp	700	
2107	Db		GACAGAAAGCT	TGGTACAAA	AGGACCTG	AATGAGATGTT	CTCCTCAGT	TTTCTCTCGG	2166	
701	Qy		ArgIleLeuLys	LeuAsnSer	ThrGluTrp	ProTyrPhe	ValValGlyIle	PheCysAla	720	
2167	Db		AGGATTCTGA	AGCTGAAC	TCAATGGCTT	ATTTTGGTGGT	GATATATTTT	GTGCT	2226	
721	Qy		IleIleAsnGly	GlyLeuGln	ProAlaPhe	SerIleIlePhe	SerArgIleIleGly	Ile	740	
2227	Db		ATTATAAACG	GGAGCCTG	CACAGCATTT	CAATAATATTT	CAAGGATTA	TATAGGATC	2286	
741	Qy		PheThrArg	AspGlu	AspPro	GluThrLys	ArgGlnAsn	SerAsnMetPhe	SerValLeu	760
2287	Db		TTTACC	CGAGATG	AGGATCTG	TAAACAAA	ACACACAGAA	TAGTAACATG	TTTCTGTATG	2346
761	Qy		PheLeuVal	LeuGlyIle	IleSerPhe	IleThrPhe	PheLeuGlnGly	PheThrPheGly	780	
2347	Db		TTTCTAGT	CTCTGGAAT	TATTTCTTTAT	TATCATTTT	TCTCCAGG	GGCTTCACATTTGGC	2406	
781	Qy		LysAlaGly	GluIleLeu	ThrLysArg	LeuArgTyr	MetValPhe	ArgSerMetLeu	Arg	800
2407	Db		AAACGCTGG	GAGATCCT	CACTAAGCGCT	TCGATACATG	TGTTTCAGAT	CCATCCTCAG	A	2466
801	Qy		GlnAspVal	SerTrpPhe	AspAspPro	LysAsnThr	ThrGlyAla	LeuThrThrArg	Leu	820
2467	Db		CAGATGCTCA	CGCTGGT	TGATGAC	CCCTAAAA	ACACCCAT	TGGAGCATG	GACACCCAGCGCT	2526
821	Qy		AlaAsnAsp	AlaAlaGln	ValLysGly	AlaIleGly	SerArgLeu	AlaValIleThr	Gln	840
2527	Db		GCCAAATG	ATGGGCTCA	AGTTAAAGGG	CTATAGGTTCC	AGGCTTGCTG	TTCATTA	ACCCAG	2586
841	Qy		AsnIleAla	AsnLeuGly	ThrGlyIleIleIle	IleSerLeu	IleTyrGly	TrpGlnLeuThr		860
2587	Db		AATATAGCA	AATCTGG	ACAGCATAT	TATATCTCT	TAATCTAT	GTGGTGG	CAATTAACA	2646
861	Qy		LeuLeuLeuLeu	AlaIleVal	ProIleIleAla	IleAlaGly	ValValIleGlu	MetLysMet		880
2647	Db		CTTTTACT	CTTAGCAAT	TGTACCCAT	CATTCGAAT	CAGGAGTGT	TGTTGAAAT	AAAAATG	2706
881	Qy		LeuSerGly	GlnAlaLeu	LysAspLysLys	GluLeuGlu	GlyAlaGlyLys	IleAlaThr		900
2707	Db		TTGTCTG	GACAAG	CATGGAAG	ATAAGAAG	AGCTAG	AAAGAGCT	GGGAGATTGCTACA	2766
901	Qy		GluAlaIle	GluAsnPhe	ArgThrVal	ValSerLeu	ThrArgGlu	ClnLysPheGluTyr		920
2767	Db		GAAGCAAT	CGAAA	ACTTTCCG	AACTGTTCTT	TTCAC	TGCGGAC	CAGAACTTTGAATAC	2826
921	Qy		MetTyrAla	GlnSerLeu	GlnValPro	TyrArgAsn	SerLeuArgLys	AlaHisIlePhe		940
2827	Db		ATGTATG	CACAGAG	TTTGC	AAGTACCA	TACAGAA	AACTCTTT	CAGGAAAGACACATCTTC	2886
941	Qy		GlyValSer	PheSerIle	ThrGlnAla	MetMetTyr	PheSerTyr	AlaGlyCysPhe	Arg	960
2887	Db		GGGGTCTCA	TTTTCTAT	CACCCAGG	CAATGAT	TATTTTCT	TATGCTGG	CTGTGTTCCGG	2946
961	Qy		PheGlyAla	TyrLeuValAla	AsnGluPhe	MetAsnPhe	GlnAspVal	LeuLeuValPhe		980
2947	Db		TTTGGT	GCTACT	TTGGTGG	CAATGAG	TTCATG	AACTTT	CAGGATGTTCTTTGCTATTC	3006
981	Qy		SerAlaIle	ValPheGly	AlaMetAla	ValGlnVal	SerSerPhe	AlaProAspTyr		1000

Db 3007 TCAGCTATTGCTTTGGTGGCCATGGCAGTGGGCGAGTCAGTTCATTGCTGCTGACTAT 3066

Qy 1001 AlAlYsAlaLysValSerAlaAlaHisValleMetileleGluLysSerProLeuIle 1020

Db 3067 GCCAAAGCCAAAGTATCAGAGCCCGCTCATCATGATCATTTGAAAGAGCCCTCGATT 3126

Qy 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040

Db 3127 GACAGCTACAGCCCTCAGCGCCCAAGCCAAATACGTTGGAAGAAATGTGACATTAAAT 3186

Qy 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060

Db 3187 GAGTGTGTTCAACTATCCCATCGCAGCAGACATCCCGTGTCTCCAGGGGCTGAGCCTC 3246

Qy 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerClyCysGlyLysSerThr 1080

Db 3247 GAGGTGAAGAGGCCAGAGCGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGAGCACA 3306

Qy 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100

Db 3307 GTTGTTCAGCTCTAGAGCGCTTCTATGACCCCTTGGCTGGTTAGTGTCTAATTCATGGC 3366

Qy 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120

Db 3367 AAAGAGATAAGCACCTGAATGTCCAGTGGCTCCGAGCAGCACCTGGGCATGCTGTCTCAG 3426

Qy 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140

Db 3427 GAGCCCATCTGTTTACGTGCAGCATTCGCGAGACATTCCTATGGAGACAAGCCGG 3486

Qy 1141 ValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisPheIle 1160

Db 3487 GTCGTATCATCAAGAGAGATTATGAGGCGAGCCAAAGGAGGCAACATACACCTTCATC 3546

Qy 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180

Db 3547 GAGACACTCCCTGAGAAATACACACAGAGTAGGAGACAAAGGAAAGCCAGCTCTCGT 3606

Qy 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200

Db 3607 GCCCAAGAACAGCGCATTTGCCATAGCTCGCGCTCTTGTGTAGACAGCTCATATTGCTT 3666

Qy 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220

Db 3667 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTGTCCAAGAGCCCTG 3726

Qy 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240

Db 3727 GACAAAGCCAGAGAGCGCGCACCTGCATTGTGATCCGCCACCGCTGTCCACCATCCAG 3786

Qy 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260

Db 3787 ATTCAGATTAAATAGTGGTGTTCAGAAATGGCAAGTCAAGAGCATGSCACACATCAA 3846

Qy 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280

Db 3847 CAGCTGTGGCTCAGAAAGCATCTATTTTCCATGATCAGTGTCCAGGCTGGAGCAAG 3906

Qy 1281 Arg 1281

Db 3907 CGC 3909

RESULT 8

CFA419568

LOCUS

DEFINITION

Canis familiaris mRNA for abortive P-glycoprotein (p-gp gene).

ACCESSION

AJ419568

VERSION

AJ419568.1 GI:17385398

KEYWORDS

P-glycoprotein; p-gp gene.

SOURCE

dog.

ORGANISM

Canis familiaris

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE

1 (sites)

AUTHORS

Roulet A.

TITLE

Characterization of a MDR1a P-gp deficient dog in relation to ivermectin sensitivity

JOURNAL

Thesis (2002) Department of Biological Sciences, Universite de Toulouse (UPS III), Toulouse, France

REFERENCE

2 (bases 1 to 3934)

AUTHORS

Roulet A.

TITLE

Direct Submission

JOURNAL

Submitted (07-NOV-2001) Roulet A., Pharmacology, Inra, 180 Chemin de Tournefeuille, 31931 Toulouse cedex, FRANCE

FEATURES

Location/Qualifiers

1..3934

gene

7..282

CDS

7..282

gene="p-gp"

7..282

gene="p-gp"

note="abortive product"

codon_start=1

evidence=experimental

product="P-glycoprotein"

protein_id="CAD12806.1"

db_xref="GI:17385398"

translation="MDPEGGRGSAEKNFWKMGKKSKKNKKKKKPTVSTFAMFRYSN WLDRLYLVTGMAALIHGAALPLMLVFGNLTALOMQEFQETKLFL"

BASE COUNT

1144 a 815 c 964 g 1011 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3934

Score: 6403.50 Matches: 1271

Percent Similarity: 99.53% Conservative: 4

Best Local Similarity: 99.22% Mismatches: 5

Query Match: 98.93% Indels: 2

DB: 4 Gaps: 1

US-09-672-725C-27 (1-1281) x CFA419568 (1-3934)

Qy 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20

Db 7 ATGGATCTCTGAAGAGGCGCTAGGGGAGTGCAGAGAGAACTTCTGGAAGATGGGCAA 66

Qy 21 LysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40

Db 67 AAAAGTAAATAAATGAGAGAAAGGAAAGAACCAACTGTGCAGCAGCTTTCGAATGTTT 126

Qy 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 60

Db 127 CGCTATTCAAATTCGCTTGATAGTTTGTATATGTTGGTGGGACAATGGCTGCCATCATC 186

Qy 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80

Db 187 CATGGAGCTGCACCTCCCTCTCATGATGCTGGTTTTTGGAAACAT----GACAGCTTGCA 242

Qy 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100

Db 243 AATGCAGGAATTTCAAGAAACAAACTTTTCCAGTTATTAATTAATGAAGAGTATTACGAAC 302

Qy 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 120

Db 303 AATACAAACATTTCACTCAACCATCTGGAGGAGAAATGACCACGTATGCCATTATTATAC 362

Qy 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140

Db 363 AGTGGGATCGTCTGGCGTGTGGTGGCTGCTTACATCCAGGTTTCATTCTGGTGGCTG 422

Qy 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160

Db 423 GCAGCAGGAGACAGATACATACTCAAAATTTAGAAAAACAATTTTTTCATGCTATCATCGCAG 482

Thu Nov 7 09:13:23 2002

QY	161	GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp	180	QY	521	LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg	540
Db	483	GAGATTGGCTGGTTGACGTGATGACCTGGTGGGAGCTTAACACCCGGCTCACAGACGAT	542	Db	1563	AAATTTGACACTCGTTGGAGAGAGAGGGCCCGCCAGCTGAGTGGTGACAGAAACAGAGA	1622
QY	181	ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIleAla	200	QY	541	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	560
Db	543	GTCTCCAAATCAATGAAGAAATCGCGACAAAATTTGAATGTCTTTCAATCAATAGCA	602	Db	1623	ATCGCAATTGCTCGGGCCCTGGTTCGCAACCCCAAGATTTCTTCGTCTGGATGAGGCAACG	1682
QY	201	ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle	220	QY	561	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	580
Db	603	ACATTTTTCACCGGTTTTATAGTGGGTTTACACGTGGTGGAGCTAACCCCTTGTGATT	662	Db	1683	TCAGCTCTCGGACACTGAAGTGAACCACTGCTCAGGTGGCCCTGGATAAGCCAGAAAA	1742
QY	221	LeuAlaIleSerProValLeuGlyLeuSerAlaIleTrpAlaLysIleLeuSerSer	240	QY	581	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	600
Db	663	TTGGCCATCAGCCCTGTCTTGGACTTTCAGCCGCATCTGGGCAAGATACTATCTTCA	722	Db	1743	GGCCGACTACCAATTGCTAGTCTCATCTGTTGTCTACAGTTCGTAATGCGGATGTCAAT	1802
QY	241	PheThrAspLysGluLeuLeuAlaLysAlaLysAlaGlyAlaValAlaGluValLeu	260	QY	601	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	620
Db	723	TTTACTGATAAAGAACTCTGGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGCTTCA	782	Db	1803	GCTGGTTTGTATGATGGAGTCAATTGTGGAGAAGGAATCATGATGAACATCATGAAAGAG	1862
QY	261	AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn	280	QY	621	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	640
Db	783	GCACCAATCAGAACTGTGATTGCTTTGGAGGACAAAAGAAAGAACTTGAAAGGTACAA	842	Db	1863	AAGGCACTTACTTCAAACTTGTCAATGCAGACAGAGGAAATGAAATTTGAGTTAGAA	1922
QY	281	LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer	300	QY	641	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	660
Db	843	AAAAATTTAGAAGAACTTAAAGAAATTTGGATTAAGAAAGCTATCACGGCCCACTTCT	902	Db	1923	AATGCCACTGGTGAATCCAAAAAGTGAAGTGAATGCTTGGAAATGCTCCAAAAGATTCA	1982
QY	301	IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr	320	QY	661	GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln	680
Db	903	ATTGGTGGCGCTTCTTATTGATCATGCAATCATATGCTCTGGCTTCTGGTATGGAGCC	962	Db	1983	GGGTCCAGTTTATAAAAAAGAGATCAACTCGCAGGAGTATACATGCCACCAAGGCCAA	2042
QY	321	SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal	340	QY	681	AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPheTrp	700
Db	963	TCCTTGGTCCCTCCAGTGAATATCTATTGACAGACTACTCTGCTCTCTTTCTGTA	1022	Db	2043	GACAGAAGCTGGTCAAAAAGAGGACTTGAATGAGAAATGACCTCCAGTTTCTCTCTGG	2102
QY	341	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	360	QY	701	ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla	720
Db	1023	TTAAATTTGGGGCTTTAGTATTGGACAGGCATCCCAAGCATTTGACATTTGCAACGCA	1082	Db	2103	AGGATTCCTGAAGCTGAACCTCAACTGAATGGCCCTATTTTGTGGTGTATATTGTGCT	2162
QY	361	ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	380	QY	721	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	740
Db	1083	AGAGGAGCAGCTTATGAAATCTTCAAGATAATTGACATAAACAAGCATTTGACAGCTAT	1142	Db	2163	ATTATAACGGAGGCGCTGCAACACAGCATTTTCAATAATATTTTCAAGGATTATAGGATC	2222
QY	381	SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	400	QY	741	PheThrArgAspGluAspProGluThrLysArgLeuArgTyrMetValPheArgSerMetLeu	760
Db	1143	TCGAAAGAGTGGACATTAACCAAGATAATTTAAGGGAATTTGGAAATTTGGAATTTCA	1202	Db	2223	TTTACCCGAGATGAGGATCTCTGAAACAAACAGCAGAAATAGTAACATGTTTCTGTATTG	2282
QY	401	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	420	QY	761	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	780
Db	1203	TTTCAAGTACCTTCTCGAAGAAGCTTAAGACCTTAAAGGCTCTCAACCTGAAGGTTTCA	1262	Db	2283	TTTCTAGTCTTGGAAATTTCTTTTATTACATTTTCTCCAGGGCTTCACATTTTGGC	2342
QY	421	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln	440	QY	781	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	800
Db	1263	AGTGGCAGACAGTGGCGCTTGGGAACAGTGGCTCGGGAAGAGCAGCCGCTGAC	1322	Db	2343	AAAGCTGGGAGATCCTCCTAAGCGGCTTCATACATGGTTTTCAGATCCATCCATCTGAGA	2402
QY	441	LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	460	QY	801	GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	820
Db	1323	CTGATGCAGAGGCTCTATGACCCACAGATGGCATGGTCTGTATTGATGGACAGACATT	1382	Db	2403	CAGGATGTCAGCTGCTTTGATGACCTTAAACACACCACTGGAGCATTCACACAGGCTT	2462
QY	461	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	480	QY	821	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	840
Db	1383	AGGACCAATAATGTAAGGCATCTCGGGAATTTACTGGTGGTGGTGGAGCCGCTGTG	1442	Db	2463	GCCAATGATGCGGCTCAAGTTAAAGGGCTATAGTTCAGGCTTCAGGCTTGTGTATACCCAG	2522
QY	481	LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	500	QY	841	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr	860
Db	1443	TTGTTTGGCCACCATGATGTAACATTCGCTATGGCCGCAAAATGTCCACATGGAT	1502	Db	2523	AAATATAGCAAACTCTGGGACAGCATTTATATCTCTTAATCTATGTTGCAATTAACA	2582
QY	501	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	520	QY	861	LeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLysMet	880
Db	1503	GAGATTGGAAGAGCTGTTAAGGAAGCAATGCCTATGATTTTATCATGAAACTACCTAAT	1562	Db	2583	CTTTTACTCTTAGCAATTTGACCCATCATTCGAATAGCAGGAGTTGTTGAATGAAATG	2642
				QY	881	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr	900

```
Db 2643 TTGCTGGCAAGCACTAAAGATAGAAAGAGCTAGAGGAGCTGGGAAGATTGCTACA 2702
QY 901 GluAlaIleGluAspPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2703 GAAGCCATCGAAACTCCGAACTGTTGTTCTTTGATCGGGAGCAGAACTTTGAATAC 2762
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2763 ATGTATGCACAGAGTTTCAAGTACCATACAGAACTCTTTGAGGAAGCACACATCTTC 2822
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2823 GGGGTCTCTATCTATCACCAGCAATGATATTTTCTATCTGCTGGCTGTTCCGG 2882
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
Db 2883 TTTGGTGCTACTTGGTGGCAATGAATTCATGAACTTTCAGGATGTTCTTTTGGTATTC 2942
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerPheAlaProAspTyr 1000
Db 2943 TCAGCTATGCTTTGGTGCCATGCGCATGGGCGAGGTCACTTCTATTTGCTTCCAGCTAT 3002
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleLeuLysSerProLeuIle 1020
Db 3003 GCCAAAGCCAAAGTATCAGCAGCCACGTCATCATGATCATTTGAAAGAAAGCCCTCTGATT 3062
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3063 GACAGCTACAGCCCTCAGGGCTCAAGCCAAATACGTTGGAAGAAATGTGACATTTAAT 3122
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3123 GAGTCGNGTTCACTATCCACTCGACACATCCCGTCTCCAGGGGCTCAGGCTC 3182
QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
Db 3183 GAGGTGAAGAAAGGCCAGACGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGACACA 3242
QY 1081 ValValGlnLeuLeuGluAtrPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
Db 3243 GTTGTTCAGCTCCFAGAGCGCTTCTATGACCCCTTGGCTGGTTCAGTGCTAATTTGATGGC 3302
QY 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120
Db 3303 AAAGAGATAAGCACCTGAATGTCAGTGGCTCGAGCACACCTGGGCATCGTGTCTCAG 3362
QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
Db 3363 GAGCCCATCTCTGTTGACTGCAGCATTTGCCGAGAACATTGCCATGGAGACAACAGCCGG 3422
QY 1141 ValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1160
Db 3423 GTCGTATCATGATGAAGATATTAGCGAGCCCAAGAGGCCAACATACACCACTTCATC 3482
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
Db 3483 GAGACACTCCCTGAGAAATACACACACAGATAGGAGACAAGGAACCCAGCTCTCTGGT 3542
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
Db 3543 GGCCAGAAACAGCGCATTTGCCATAGCTCGCGCTCTTGTAGACAGCCCTCATATTTGCTT 3602
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
Db 3603 TTGGATGAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAGGCCCTG 3662
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
Db 3663 GACAAAGCCAGAGAGCCGACCTGCTGATTCGATTCGCCACCGCTGTGCCACCATCCAG 3722
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
```

```
Db 3723 AATCAGATTAAATAGTGGTGTTCAGAAATGCAAGTCAAGGATGCGCACACATCAA 3782
QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
Db 3783 CAGTGTGCTGCTCAGAAAGGCATCTATTTTCCATGATCAGTGTCCAGGCTGGAGCAAA 3842
QY 1281 Arg 1281
Db 3843 CGC 3845
RESULT 9
LOCUS AX322787 3860 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 1 from Patent WO0192877.
ACCESSION AX322787
VERSION AX322787.1 GI:18093766
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Sorrentino,B. and Schuetz,J.
TITLE Method of identifying and/or isolating stem cells
JOURNAL Patent: WO 0192877-A 1 06-DEC-2001.
FEATURES
location/Qualifiers
source
1..3860
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1135 a 746 c 957 g 1022 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3860
Score: 5866.00 Matches: 1164
Percent Similarity: 95.09% Conservative: 55
Best Local Similarity: 90.80% Mismatches: 59
Query Match: 90.62% Indels: 4
DB: 6 Gaps: 4
US-09-672-725c-27 (1-1281) x AX322787 (1-3860)
QY 1 MetAspProGluGlyGlyArgLysGlySerAla--GluLysasnPheThrLysMetGly 19
Db 1 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAAAGAACTTTTAACTGAAC 60
QY 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 61 AATAAAAGT--GAAAGATAGAGAGAAAGAAACCAACTGTCAGTGTATTTCAATG 117
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
Db 118 TTTTCGCTATTCAAATTTGGCTTGACAAAGTTGTATATGGTGGTGGAACTTTGGCTGCCATC 177
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 178 ATCCATGGGGCTGGACCTCTCTCATGATGCTGGTGTGGAGAAATACAGATATCTTT 237
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 238 GCAATATGCAGGA--AATTTAGAAGATCTGATGTCAAAACATCCTAATAGAAGTGTATC 294
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 119
Db 295 AATGATACAGGGTCTTCTCATGAAT--CTGGAGAGAGACATGACCAGATATGCTTATTAT 351
QY 120 TyrSerGlyIleGlyValLeuValAlaTyrIleGlnValSerPheTrpCys 139
Db 352 TACAGTGGAAATTTGGTCTGGGGTCTGGTGTCTGTCTTACATTCAGGTTTCATTTGGTGC 411
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
```

Db 412 CTGGCAGCTGGAAGACAAATACACAAAATAGAAAAACAGTTTTTTCATGCTATAATGCGA 471
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 472 CAGGAGATAGGCTGGTTGATGTGCACGATGTGGGGAGCTTAACACCGCAGCTTACAGAT 531
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db 532 GATGCTCTAAGATTAAATGAAGGATTGTGTGACAAAATTTGGAATGTTCTTTCAGTCAATG 591
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrVal 219
Db 592 GCAACATTTTTCACCTGGTTTATAGTAGGATTTACACGTGGTTGGAAGCTAAACCCCTGTG 651
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 652 ATTTTGGCCATCAGTCCTGTTCTGTGACGTGCAGCTGCTGTGGCAAGATATATCT 711
Qy 240 SerPheThrAspLysGluLeuAlaIleValAlaLysAlaGlyAlaValAlaGluVal 259
Db 712 TCATTTACTGNTAAGACTCTTAGCGTATGCAAAAGCTGGAGCAGTACCTGAAGAGGTC 771
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyr 279
Db 772 TTGGCAGCAATTAGAATCTGATGCTGATTTGGAGGACAAAAGAAAGCTTGAAGGTTAC 831
Qy 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 832 ACAAATAATTTAGAAGAGCTAAAGAAATTTGGATTTGAAGAAAGCTATTTACGCCAATAT 891
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 892 TCTATAGTCTGCTTTCCTGCTGATCTATGTCATCTTATGCTCTGGCCCTTCTGGTATGG 951
Qy 320 ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPheSer 339
Db 952 ACCACCTTGGTCTCTCAGGGGAATATCTATTGGCAAGTACTCACTGATTTCTTTTCT 1011
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1012 GTATTAATGGGGCTTTTAGTGTGTGACAGGCATCTCCAAAGCATTTGAAGCATTTGCCAAT 1071
Qy 360 AlaArgGlyAlaAlaIleValGlyIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1072 GCAAGAGGACGCTATGAAATCTTCAAGATAATTTGATAATTAAGCAAGTATTGACAGC 1131
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1132 TATTCGAAGAGTGGGCACAAACACAGATAATTTAAGGGAAATTTGGAATTCAGAAATGTT 1191
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1192 CACTTCAGTTACCCATCTCAAAAGAGTTAAGATCTTGAAGGCGCTGAACCTGAAGGTG 1251
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrVal 439
Db 1252 CAGAGTGGGCAGACGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGAGACACACAGTC 1311
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1312 CAGCTGATGACAGGGCTCTATGACCCACAGAGGGGATGTGCTAGTGTGATGGACAGGAT 1371
Qy 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1372 ATTAGGACCATAAATGTAAAGGTTTTCACGGGAATTCATTGGTGGTGAGTCAGGAACCT 1431
Qy 480 ValLeuPheAlaThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1432 GTATTGTTGGCCACGACGATAGCTGAAACATCTCGCTATGGCGGTGAAAAATGTCAACATG 1491
Qy 500 AspGluIleGlyAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1492 GATGAGATTGAGAAAGCTGTCAAGAAAGCCATTCCTATGACCTTATCATGAACATGCGCT 1551

Qy 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1552 CATAAATTTGACACCCCTGGTTGGAGAGAGAGGGGCCAGTTGAGTGTGGGCAGAGCAG 1611
Qy 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 1612 AGGATCGCAATTGCACGTGCCCTGGTTGCGAACCCCAAGATCTCTCTGCTGGATGAGGCC 1671
Qy 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 1672 ACGTCAGCCTTTGGACACAGAAAGCGAAGCAGTGTTCAGGTGGCTCTGGATAAGCCAGA 1731
Qy 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 1732 AAAGTCGACACCACTTGTGATAGCTCATCGTTTGTCTACAGTTCGTAATGCTGACGTC 1791
Qy 600 IleAlaGlyPheAspAspGlyValIleValGlnLysGlyAsnHisAspGluLeuMetLys 619
Db 1792 ATCGCTGGTTTCGATGATGGAGTCATTGTGGAGAAAGAAATCATGATGAACCTCATGAA 1851
Qy 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 1852 GAGAAAGCATTACTTCAAACTTGTCAATGCGACAGCAGCAAAATGAAGTTGAATTA 1911
Qy 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 1912 GAAATGCGAGCTGATGAATCCAAAATGAAATGATGCTTGGAAATGCTTCAAATGAT 1971
Qy 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 1972 TCAAGATCCAGTCTAATAAGAAAAGATCAACTCGTAGAGTGGCCGTGGATACACAGCC 2031
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2032 CAAGACAAAGCTTAGTACCAAGAGCGCTCTGGATGAAGATATACCTCCAGTTTCCCTT 2091
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2092 TGGAGGATTATGAAGCTAAATTAACCTGAATGGCCCTATTTTGTGTGGTATTGTGT 2151
Qy 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2152 GCCATTAAATGGAGCGCTGCACACGACATTTGCAATAATATTTTCAAGATATAGGG 2211
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2212 GTTTTACAGAAATTGATGATCCTGAAACAAAACGACAGAATAGTAACCTGTTTTCAC 2271
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2272 TTGTTTCTAGCCCTTGGAAATTATTTCTTTTATACATTTTTCCTTCAAGGTTTTCACAT 2331
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2332 GCGAAGCTGGAGAGATCCTCACCAAGCGGCTCCGATACATGGTTTCCGATCCATGCTC 2391
Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2392 ACACAGGATGTGAGTTGGTTTGTGATGACCTTAAACACACCACTGGAGCATTGACTACC 2451
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2452 CTCGCAATGATGCTCAAGTTAAAGGGGCTATAGTTCAGGCTTGCCTGTGTAATTACC 2511
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2512 CAGAATATAGCAAAATCTTGGGACAGGAATAATATATCTTCATCTATGTTGGCAACTA 2571
Qy 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2572 ACACCTGTTACTCTTAGCAATTTGACCCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAA 2631

```
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
|||||
Db 2632 ATGTTGTCGACAGCACTGAAGATGAAGAAAGCACTGAAGGTCGCGGAAGATCGCT 2691
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
|||||
Db 2692 ACTGAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTGACTCAGGAGCAAGAGTTGAA 2751
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
|||||
Db 2752 CATATGATGTCAGAGTTTCAGGTACCATACAGAACTCTTTGAGAAAGCACACATC 2811
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 2812 TTTGGAATTACATTTCTTACCAGCAATGATGATATTTTCTTATGCTGGATGTTTC 2871
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
|||||
Db 2872 CGGTTTGGAGCCTACTTGGTGGCATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 2931
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
|||||
Db 2932 TTTTCAGCTGTGTCCTTTGGTGCCATGCGCGTGGGCAAGTCAGTTTATTGCTCCCTGAC 2991
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleLeuGluLysSerProLeu 1019
|||||
Db 2992 TATGCCAAGCCAAAATATCAGCGCCACATCATCATGATCATTTGAAAACCCCTTGG 3051
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
|||||
Db 3052 ATTGACAGCTACAGCAGGAAGCGCTAATGCCAACACATTTGGAAGAAATGTCCACATTT 3111
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
|||||
Db 3112 GGTGAAGTTGTATTCAACTATCCACCGCGACCGACATCCAGTCTTCAGGGACTCAGC 3171
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
|||||
Db 3172 CTGGAGGTGAAGAGGCCAGACCGCTGGCTCTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3231
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
|||||
Db 3232 ACAGTGTGTCCAGCTCTCTGGAGCGTTCTACGACCCCTTGGCAGGAAAGTGTCTGTGAT 3291
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
|||||
Db 3292 GCGAAGAAATAAGCGCTGAATGTTTCAGTGGCTCCGAGCACACCTGGGCGATGCTGCC 3351
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
|||||
Db 3352 CAGGAGCCCATCTTGTGACTGCAGCATTTGCTGAGAACATTTGCTATGGAGACACAGC 3411
QY 1140 ArgValValSerHisGluGluIleValGlnAlaLysGluAlaAsnIleHisHisPhe 1159
|||||
Db 3412 CGGTTGTGTCCAGGAAGAGATCGTGGGCGCAGCAAGAGGCGCAACATACATGCTTC 3471
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
|||||
Db 3472 ATCCAGTCACTGCCCTATAATAATATAGCACTTAAATGAGGACAAAGGAAGTCACTCTCT 3531
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
|||||
Db 3532 GGTGCCAGAAACCAACGATGCGCATAGCTGCGCCCTTGTAGACAGCCCTCATATTTTG 3591
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
|||||
Db 3592 CTTTTGGATGAAGCCAGCTCAGCTCTGCATACAGAAAGTGAAAGAGTTGTCCAAAGAGCC 3651
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 3652 CTGACAAAGCCAGAGAGCGCCGACCTGCTGATTTGCTCACCAGCCCTGTCACCATC 3711
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||

Db 3712 CAGATGCAGACTAATAGTGGTGTTCAGAAATGGCAGATCAAGAGCATGCGACGCAT 3771
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 3772 CAGCAGCTGTGCGACAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 3831
QY 1280 LysArg 1281
|||||
Db 3832 AAGCGC 3837
RESULT 10
LOCUS E02326
DEFINITION Multidrug resistance relating gene derived from human normal cells.
ACCESSION E02326
VERSION E02326.1 GI:2170561
KEYWORDS JP 1990100680-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4378)
AUTHORS Ueda,K. and Komano,T.
TITLE HUMAN NORMAL CELL-DERIVED MDR RELATED GENE
JOURNAL Patent: JP 1990100680-A 1 12-APR-1990;
SUNTORY LTD
COMMENT OS Homo sapiens
PN JP 1990100680-A/1
PD 12-APR-1990
PF 05-OCT-1988 JP 1988251475
PI UEDA KAZUHIRO, KOMANO TORU
PC C12N15/12,C12N1/21,C12Q1/68;
CC strandedness: Single;
CC topology: Linear;
CC *source: tissue type=Adrenal gland;
CC *source: clone=SAM132;
FH Key Location/Qualifiers
FT 5'UTR 1..137
FT CDS 138..3980
FT /gene='Multidrug resistance relating gene' FT
FT mat_peptide 138..3977
FT /gene='Multidrug resistance relating gene' FT
FT 3'UTR /gene='MDR1'
FT Location/Qualifiers
1..4378
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 1315 a 818 c 1057 g 1188 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4378
Score: 5863.00 Matches: 1163
Percent Similarity: 95.09% Conservative: 56
Best Local Similarity: 90.72% Mismatches: 59
Query Match: 90.58% Indels: 4
DB: Gaps: 4
US-09-672-725C-27 (1-1281) x E02326 (1-4378)
QY 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheThrLysMetGly 19
|||||
Db 138 ATGGATCTTTGAAGGGGACCGCAATGGAGGAGCAAGAAAGAACTTTTAACTGAAC 197
QY 20 LysLysSerLysLysLysLysLysGluLysLysLysProThrValSerThrPheAlaMet 39
|||||
Db 198 AATAAAAGT---GAAAGAGATAAGAGGAAAAAGAACCAACTGTGAGTGTATTTCAATG 254
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
|||||
```


255	Db	TTTCGGCTATTCAAATTTGGCTTGACAAAGTTGTATATATGGTGGGAAACTTTGGCTGGCAATC	314
60	Qy	IleHisGlyAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe	79
315	Db	ATCCATGGGGCTGGACTTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT	374
80	Qy	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleLeuAsnGluSerIleThr	99
375	Db	GCAATGTCAGGA---AATTTAGGAAGATCGATGTGCAAAACATCACTAATAAGAGTGATATC	431
100	Qy	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr	119
432	Db	AATGATACAGGGTCTTCATGAAT--CTGGAGGAAAGACATGCACCGGTATGCCATTAT	488
120	Qy	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys	139
489	Db	TACAGTGGAAATGGTGCTGGGGTGTGGTTCGCTTCACATTCAGGTTTCATTTGGTGC	548
140	Qy	LeuAlaIaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159
549	Db	CTGCGACGCTGGAGACAAATACACAAAATTAGAAAACAGTTTTCATGCTATAATGCGA	608
160	Qy	GlnGluIleGlyTrpPheaspValHisaspValGlyGluLeuAsnThrArgLeuThrAsp	179
609	Db	CAGGAGATAGCTGGTTGATGTGCACGATGTGGGGAGCTTAACACCCGACTTACAGAT	668
180	Qy	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle	199
669	Db	GATGTCCTCAAGATTATGAAGGAATTGGTGACAAAATTGGAATGTCTTTCAGTCAATG	728
200	Qy	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219
729	Db	GCAACATTTTCACTGGGTTTATGTAGGATTTACACGTGGTGAAGCTAACCCTTGTG	788
220	Qy	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239
789	Db	ATTTTGGCCATCAGTCGTCTTGGCAGTGTGCAGCTGCTGCTGGGCAAAAGACTATCT	848
240	Qy	SerPheThrAspLysGluIleuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal	259
849	Db	TCATTTACTGATAAAGACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC	908
260	Qy	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr	279
909	Db	TTGCGACCAATTAGAACTGTGATTCATTTGGAGGACAAAAGAAAGACTTGAAAGGTAC	968
280	Qy	AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299
969	Db	AACAAAAATTTAGAAGAAGCTAAAGAAATGGGGTAAAGAAGCTATTACAGCCAATATT	1028
300	Qy	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319
1029	Db	TCATATAGGTGCTCTTCCTCGTGAATCATGATCTTATGCTCTGGCCCTCTCGTATGGG	1088
320	Qy	ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer	339
1089	Db	ACCACCTTGGTCTCTCAGGGGAATATTCTATTGGCAAGTACTCACTGTATCTTTTCT	1148
340	Qy	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359
1149	Db	GTATTAATTGGGGCTTTTACTGTGGCAGGCATCTCCAAGCATTTGAAGCATTTGCAAAAT	1208
360	Qy	AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379
1209	Db	GCAAGAGGACGACTATGAAATCTTCAAGATAATTGATAATAAGCCAAGTATTGCAGC	1268
380	Qy	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399
1269	Db	TATTCGAAGAGTGGGCACAAACCGATATATATAGGGGAATTTGGAATTCAGAAATGTT	1328
400	Qy	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419
1329	Db	CACTTTCAGTTACCCATCTCGAAAGAAGTTAAGATCTTTGAAGGGTCTGAACCTGAAGGTG	1388


```
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2469 GGCAGAGCTGGAGAGATCCTCACCAAGCGCTCCGATACATGGTTTCCGATCATGCTC 2528
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2529 AGACAGGATGTGAGTTGGTTTGTGATGACCTAAAAACACCACTGGAGCATTGACTACCAGG 2588
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2589 CTCGCCAATGATGTCTCAAGATTAAAGGGCTATAGGTTCCAGGCTTGCTGTAAATACC 2648
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2649 CAGAATATAGCAAACTTTGGGACAGGAATAATATATCTTCATCTATGGTTGGCACTA 2708
QY 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2709 ACCTGTTTACTCTTTAGCAATTTGTACCCATCATTCGCAATAGCAGGAGTTGTTGAAATGAA 2768
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2769 ATGTTGTCGACAGCACTGAAAGATAAGAAAGACTAGAAAGTCTCTGGAGATCGCT 2828
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2829 ACTGAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAGATTGAA 2888
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2889 CATATGTATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTAGGAAAGCACACATC 2948
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2949 TTTCGAATATACATTTCTTCACCCAGGCAATGATGATTTTCTTCATGCTGGATGTTTC 3008
QY 960 ArgPheGlyAlaTyrLeuValAlaAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 3009 CGGTTTGAGCCTACTTGGTGGCACAATAAATCATGAGCTTTGAGGATGTTCTGTAGTA 3068
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3069 TTTCACGCTGTGCTTTGTGGCATGCGCGTGGGCAAGTCAGTTCATTTGCTCCTGAC 3128
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3129 TATGCCAAAGCCAAATATCAGCAGCCACCATCATCATGATCATTTGAAAAACCCCTTTG 3188
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3189 ATTGACAGCTACAGCAGGAAGCCCTAATCCGCAACACATTCGAAGGAATGTCACATTT 3248
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3249 GGTGAAGTTGATTTCAACTATCCACCCGACCGACCATCCAGTGTCTCAGGACCTGAGC 3308
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3309 CTGGAGGTGAAGAGGCCAGACGCTGGCTCTGCTGGGACAGCTGCTGGGAGAGC 3368
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuLeuAsp 1099
Db 3369 ACAGTGTCTCAGCTCTCGGAGCGGTTCTACGACCCCTTGGCAGGAAAGTGTGCTGAT 3428
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3429 GGCAGAAATAAAGCGATGTAATGTTCAGTGGCTCCGACACACCTCGGCACTCGTGTCC 3488
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3489 CAGAGCCCACTCTGTTGCTGCTGACCATGCTGAGAACATTCGCTTCATGGAGACACAGC 3548
QY 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAlaHisPhe 1159

||||| 3549 CGGTTGGTGTACAGGAAGAGATTGAGGGCAGCAAGAGAGGCCAACATACATGCTTC 3608
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3609 ATCGAGTCACTGCCTAATAATATAGCACTAAAGTAGGAGACAAAGAACTCAGCTCTCT 3668
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3669 GGTGGCCAGAAAACACGATTCGCTAGCTCGTGCCTTTGTTAGACAGCTCATATTG 3728
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3729 CTTTTGGATGAAGCCAGCTCAGCTCGGATACAGAAAGTTGCAAGAGGCC 3788
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3789 CTGGCAAAAGCCAGAGAGCCGACCTGCATTGTGATTGCTCAGCGCTGTCACCATC 3848
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3849 CAGAATGCAGACTAATAGTGTGTTTCAGAAATGGCAGAGTCAAGGAGCATGGCAGCAT 3908
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3909 CAGCAGCTGCTGGCACAGAAAGCATCTATTTTCAATGTCAGTGTCCAGGCTGGAACA 3968

RESULT 11
AXI08654
LOCUS AXI08654
DEFINITION Sequence 1 from Patent WO0123565.
ACCESSION AXI08654
VERSION AXI08654.1 GI:13923886
KEYWORDS
SOURCE Macaca fascicularis.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (bases 1 to 4186)
Stocker,P.J., Steimel-Crespi,D.T. and Crespi,C.L.
P-glycoproteins from macaca fascicularis and uses thereof
Patent: WO 0123565-A 1 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
Location/Qualifiers
1..4186
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
100..3942
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37764.1"
/db_xref="GI:13923887"
/translation="MDLEGDRNGAERKNFFKLNKSKDKKPKTVSVFSMPFRYSN
WLDKLYMVGTALAIHAGAGLPLMLVFGMDTDTFANAGLGDIGALLTNSNITDTV
PVMLEEDMTRYAYYSYGIGAGVLAAY IQVSWCLAAAGRIQHKIKOPFHAIWROEI
GWFVDVHDELNTRLTDDVSKINBGIKGMFQSMATFTFTGTVGTRGWLTLVI
LAISPVGLSLAAVWAKILSSFTDKELLAYAKAGAAVEVLAAITVIAFGQKKELER
YKNLEAKRIGIKKAITANISGAALLIYASTALFYGTITLVLSKEYSIGVLTIV
FFSVLIGAFSVGQSPSIEAFANAGAAFEIKLIDNKPSIDSYSGKHDPKIKGNI
EFRNVHFSYPSREVKILKGLNKVQSGTVALVNSGCGKSTVQLMQRLYDPTGEM
VSDGQDITINVRFLIIGVSOEPLVATTAEINIRYGRDVTMDTEKAVTEAM
AYDFIMKLPOKEDTLVGERGALSGGOKRIATARLVNPKILLIDEATSKDITSE
AVVQVALDKARKGTTIVIAHRLSTVRNADVIAGDGGGVIVKGNHDELMKEGIIFK
LVTMQTAGNEIELENADESKEIDTLEMSHSDSLIRKRSVRSGQODRKL
STKDALDEIPPSVFRIMKLNLTWFFYVFGCAITINGLQDPAFVIFKIGFT
NRDDAETKRONSLNLSLFLVLTGTVSTFFLOGTFGKAGELTKRLRYWPSMLR
QDVSWFDDPKNTTGTALTRLANDAAOVKGAIGSLAIITONIANIGTITSLIQWQ
LTLTLLAIVPIIAGVVMKLSQALKDKLEGAKTATEAIENFRIVVSLTQEQ
```

KFERMYDOSLOVPYRNSLRKAHIFGTFITFSFTQAMMYFSYAGCFRFGAYLVVAHLSMSE
 DYLLVFSAVFEGAMVGVSSFPADYAKAKVSAHHIWIIEKTPLDISYSTGLKPNV
 LEGNITFNEVFNPTLRDIPVLOGLSLEVKKGOTLALVSGSCGCKSTVQOLLREYD
 PLAKVLDDKEIKOLINQWMLRHGIVSQEPILFDCSISIENTAGDGNRSRVSQEELV
 RAAKEANTHAPIESLPNKYSRVDKGTQLSGGQKRIATARALVRPHILLDEATS
 ALDTESEKVOEALDKAREGTRCTVIAHRLSTIQNALDILVVFQNGRVRKEHGTQOOLLA
 QKGLVFSMVSVOAGAKRQ"

BASE COUNT 1226 a 799 c 1039 g 1122 t
 ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4186
 Score: 5862.00 Matches: 1163
 Percent Similarity: 95.32% Conservative: 59
 Best Local Similarity: 90.72% Mismatches: 56
 Query Match: 90.56% Indels: 4
 DB: 6 Gaps: 4

US-09-672-725C-27 (1-1281) x AX108654 (1-4186)

QY 1 MetAspProGluGlyArgLysGlySerAlaGlu---LysAsnPheTrpLysMetGly 19
 DB 100 ATGGATCTTGAAGGGGACCGCAATGGAGGAGGAGAGAAAGAACTTTTAAACTGAAC 159
 QY 20 LysLysSerLysLysGluLysLysProThrValSerThrPheAlaMet 39
 DB 160 AATAAAGT---AAAAAGATAGAAGAAAGAAACCAACTGTCAGTGTATTTCATG 216
 QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLalle 59
 DB 217 TTTTCGCTATTCAAAATGGCTTGACAAAGTTGTATATGGTGGGAACTTTGGCTGCCATC 276
 QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
 DB 277 ATCCATGGAGCTGGACTTCTCTCATGATGCTGGTGTGGAGACATGACGATACCTTT 336
 QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
 DB 337 GCAAAATGCAGGA---AATTAGGAGATTTAGGAGCTCTGTGACTAATAGCAGTAATATC 393
 QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
 DB 394 ACTGATACAGTGCCTGCATGAAT---CTGGAGGAAGATATACCGAGTATGCTATTAT 450
 QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
 DB 451 TACAGTGGAAATGGTCTCGGGTGTCTGCTGCTTACATTCAGGTTTCATTTTGGTGC 510
 QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
 DB 511 CTGGCAGCTGGAAGACAAATACAAAAATTAGAAAAACAGTTTTTTCATGCTATAATGCCA 570
 QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
 DB 571 CAGGAGATAGCGTGGTTGTATGTCAGATGTTGGGAGCTTAACACCGGCTTACAGAT 630
 QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
 DB 631 GATGCTCCCAAGATTAAATGAAGAAATGGTGCAAAAATGGAAATGTTCTTCAGTCAATG 690
 QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
 DB 691 GCAACATTTTTCATCGGGTTTATAGTAGGATTTACACGTGGTGGAGAGCTAACCCCTGTG 750
 QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
 DB 751 ATTTTGGCCATCAGTCTCTTCTTGGACTGTCAGTGCAGTCTGGGCAAGATACCTGTCT 810
 QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
 DB 811 TCATTTACTGATAAAGAAGAACTTACGTTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 870
 QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyr 279

DB 871 TTGGCAGCAATTAGAAGCTGTGATTGGAGGACAAAGAAAGAACTCGAAAGGTAC 930
 QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle 299
 DB 931 AAAAAAATTTAGAAAGAGCTAAAGAAATTTGGGATAAAGAACTATTACAGCAATATT 990
 QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
 DB 991 TCTATAGGTGCTGCTTCCCTGCTTATCTATGTCATCTTATGCTCTGGCCTTCTGTATGG 1050
 QY 320 ThrSerLeuValLeuSerSerGlyThrIleGlyGlnValLeuThrValPheSer 339
 DB 1051 ACCACCTTGGTCTCTCAAGGAATATTCTATTGGACAACTACTACTGTATTCTTTCT 1110
 QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
 DB 1111 GTATTAATTTGGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTGAAGCATTTCGCAAT 1170
 QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
 DB 1171 GCAAGAGGAGCAGCTTTTGAATCTTCAAGATAATTGATAATTAAGCCAACTATTGACAGC 1230
 QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
 DB 1231 TATTCGAAGAGTGGGCACAAACCAAGATAATTAAGGGAATTTGGAAATTCAGAAATGTT 1290
 QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
 DB 1291 CACTTCAGTTACCATCTCGAAAGAAAGTTAAGATCTTGAAGGGCCTGAACCTGAAGGTG 1350
 QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
 DB 1351 CAGAGTGGCAGAGCGGTGGCTGTGGAAACACGCGCTGTGGGAAGAGCACAAACGGTC 1410
 QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
 DB 1411 CAGCTGATGACAGAGCTTTATGACCCACAGAGGCGATGTCAGTGTGTGAGACAGGAT 1470
 QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
 DB 1471 ATTAGGACCAATAACGTAAGCTTTCTACGGAAATCATCGGTGTGGTGAAGCAACCT 1530
 QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
 DB 1531 GTATTTGTTGCCACACACATAGCTGAAACCATTCGCTATGCTGTGAGAGATGTCACCATG 1590
 QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
 DB 1591 GATGAGATTGAGAAAGCTGTCAAGGAAGCCAAATGCTATGACTTTATCATGAAACTGCCT 1650
 QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnGln 539
 DB 1651 CAGAAATTTGACACCTGTTGGAGAGAGAGGGGCCAGCTGAGTGGTGGGCAAGAGCAG 1710
 QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
 DB 1711 AGGATCGCATTCGACGTCGCTGCTGCGAACCCCAAGATCCCTCCTGCTGGACAGGCC 1770
 QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
 DB 1771 AGCTCAGCCTTGGACACAGAAAGTGAAGCAGTGGTTCAGGTGGCTCTGGATAAGGCCAGA 1830
 QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
 DB 1831 AAAGTCGACACCACTTGTGATAGTCATGTTGCTACGGTTTCGTAATTCGCGACGTC 1890
 QY 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
 DB 1891 ATCGCTGTTTTCATGATGGAGTCAATTTGGAGAAAGAAATCATGATGATGATCATGAAA 1950
 QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639

Db 1951 GAGAAAGGCAATTACTTCAAACCTTGTCAACATGCACAGCAGGAGAAATCAAAATGAATTA 2010
Qy 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2011 GAAATGACGCTGATGAATCCAAAGTGAATGTGACCTTGGAAATGCTTTCACATGAT 2070
Qy 660 SerGlySerSerLeuLeuLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2071 TCAGGATCCAGTCTAATAAGAAAAAGATCCACTCGTAGGAGTGTCGGGATCAAGGC 2130
Qy 680 GluAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
Db 2131 CAAGACAGAAAGCTAGTACCAAGAGGCTCTGGATGAAGTATACCTCCAGTTTCCTTT 2190
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyLylePheCys 719
Db 2191 TGGAGATATGAAGCTAAATTAACCTAGTGGCCCTAATTTGTTGTTGGTGTATTTGT 2250
Qy 720 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2251 GCCATTATAAATGGAGGCTGCAACCCAGCAATTTGCAGTAATATTTTCAAGATATAGGG 2310
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2311 ATTTTTCACAGAAATGATGATGCCGAAACAAAACGACAGATAGTAATCTGTTTTCATA 2370
Qy 760 LeuPheLeuValLeuGlyIleLeuSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2371 TGTTCCTAGTCCTTGGAAATGTCTCTTTTATTAATTTTCTTCAGGGCTTCACATTT 2430
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2431 GSCAAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATGGTTTCCGATCCATGCTC 2490
Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2491 AGACAGGATGTGAGCTGTTGATGACCTTAAACACACCATCGGAGCATTCACATCCAGG 2550
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2551 CTCGCCAATGATGCTGCTCAAGTTAAGGGCTATAGTTCACGGCTTGCTTATATATACC 2610
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2611 CAGAAATATAGCAAAATCTTGGACAGGAATAATATATCTTAATCTATGTTGGCACTG 2670
Qy 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2671 ACACGTGTTACTCTTAGCAATTTGATCCCATCATTCGAATAGCAGGAGTTGTTGAAATGAA 2730
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2731 ATGTTGCTGGACAGGCACTGAAAGATGAAGAAAGCACTAGAAAGTGTGGGAATCGCT 2790
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2791 ACTGAGCAATAGAAAACCTCCGAACTGTTCTTTGACTCAGGAGCAGAAATTTGAA 2850
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2851 CATATGTATGATCAGAGTTTCAGGTACCATACAGAAATCTTTTGGAGAAAGCACATC 2910
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2911 TTTGGAATTCACGTTTCTCCACGAGCAATGATGATTTTTCCTATGCTGGATGTTTC 2970
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 2971 CGGTTTGGAGCTACTTGGTGGACATAGTCTCATGAGCTTTGAGGAGTTCTGTTAGTA 3030
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3031 TTTTTCAGCTGTGTGCTTTGGTGCCATGGCCGTGGGCAAGTCAGTTTCATTTGCTCTGAC 3090

Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3091 TATCCAAAGCAAGATATCAGCAGCCACATCATCATGATCATTTGAAAAACCCCTTTG 3150
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3151 ATTACAGCTACAGCAGACAGAGGCCTAAAGCCGAACACATTTGGAAGAAATGTCACATTT 3210
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3211 AATGAAGTTGATTTCAACTATCCACCCGAGCTGCACATCCAGTGTCTCAGGGGCTGAGC 3270
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3271 CTGGAAGTGAAGAGGCCACAGACCTGGCCCTGTGGGCAGCAGTGGCTGTGGAGAGC 3330
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3331 ACGTGTCTCCAGCTCTCTGGAGCGGTTCTATGACCCCTTGGCGGGAAGTGTCTGTGAC 3390
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3391 GGCAAGAAATAAGCAACTGAATGTTTCAGTGGCTCCGAGCACACCTGGGCATCGTCTC 3450
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3451 CAGGAGCCATCTCTGTTGACTGCAGCATTAGTGAACATTCCTATGGAGACAAACAGC 3510
Qy 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3511 CGGTGTGTGTACAGGAGAGATCGTGGAGGCGACCAAGAGGCCAATATACACGCTTC 3570
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3571 ATCGAGTCACTGCTAATAAATATAGCACCAGTAGTAGGAGACAAAGAACTCAGCTCTCT 3630
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3631 GGTGGCCAGAAACACGCAATTCGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3690
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3691 CTTTGTGATGACCCACATCAGCTCTGGATACAGAAAGTGAAAGGTTGTCCAAAGAGCC 3750
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3751 CTGGACAAAGCCAGAGAGCCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3810
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3811 CAGAATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGAGTCAAGAGACCGCACACAT 3870
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3871 CAGCAGCTGCTGGCAGAGAAAGCACTATTTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 3930
Qy 1280 LysArg 1281
Db 3931 AAGCGC 3936

RESULT 12

AX108656
LOCUS AX108656 4195 bp linear PAT 30-APR-2001
DEFINITION Sequence 3 from Patent WO0123565.
ACCESSION AX108656
VERSION AX108656.1 GI:13923888
KEYWORDS crab-eating macaque.
SOURCE Macaca fascicularis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopithecinae; Macaca.

```
REFERENCE
AUTHORS      1 (bases 1 to 4195)
TITLE        Stocker,P.J., Steimel-Crespi,D.T. and Crespi,C.L.
JOURNAL      P-glycoproteins from macaca fascicularis and uses thereof
              Patent: WO 0123565-A 3 05-APR-2001;
              GENTEST CORPORATION (US)
FEATURES
  source      Location/Qualifiers
              1..4195
              /organism="Macaca fascicularis"
              /db_xref="taxon:9541"
              100..3951
              /note="unnamed protein product"
              /codon_start=1
              /protein_id="CAC37765.1"
              /db_xref="GI:13923889"
              /translation="NDLEGDRNGAEKKNFFKLNKSKDKKERKTPVSVFSPFRYSN
              WLDKLYMVVGTLLAIIHGAGLPLMLVFGDMDFPANAGNLGDGLGALLFNNTSSNIT
              DTVPYNNLEEDMTRAYIYSGIGAGLVAAAYQSVFCLAAAGRLQHKIRKQFFRAIMR
              QBIGFVDHVDENLRLTDDSKINEGIDKIGNMFFOSMATFFTFIVGTRGKWL
              LVLAISPVGLSAAVAKIILSSFTDKELLAYAKAGAAVEVLAARIVTIAFGQKKE
              LFRYNNLEAKRIGIKKAITANISIGAAFLLIYASALAFWYGTIVLSKEYSIGOV
              LTVFSVLIGAFSVGOASPTAPANARGAPEFKIIDNKPISDSYSKSHKLPDNK
              CNLERNVHFSVPSKVEKILKGLNLKVSQGTVALVNSGCGKSTVQLMQRLDPT
              EGMVSDGDIKTIINVRFLREILIGVVSQEPVLFAITIAENIRYGRDVTMDEIEKAYK
              EANAYDFIMKLPQKEDTILVGERAQLSGGQKQRIARALVRNPKILLDLDEATSAIDT
              ESEAYVQVALDKARKGRRTIVIAHRLSTRNADVIAGDDGVIYEKGNHDELMKEGI
              YKLVMTQTAGNEILENAADESKSEIDTLEMSHDSGLIRKSTRSRVSGSQGD
              RKLSTKEALDESIPIVSWFRIMKMLNLTWPIVVGVCALINGLOPAFAVIFSKIIIG
              IETRNDDATEKONSLFLPLVLGIVSFITFELQGTFGKAGEILTKRLRYNVERS
              MLRQDVSWFDDPKNTGALTFLRLANDAAQVKGAGSRLAIIQNTIANLGTLIIISLY
              GWOLTLILAIVPIIAIAGVEMKLSGOALKDKKELEGAKIKAEALENPRTVSVSLT
              QOKEFHEMDQSLQVPRNSLRKAHIFITFTSQAMMYFSYAGCFREGAYLVASHLM
              SPEDVLVFSVAVFAMAVGQVSFPADYAKAKVAAHIIMIIEKPLIDISYSEGLK
              PNTLEGNTVLLGVEFNYPTRLDIPVLQGLSLEVKKGQTLALVSGSGCGKSVQVOLLER
              FYDPLAGKVLNDEKIKOLNQMWRHLGIVSOEPIFLDCSTISENIAYGDNRSVVSOE
              EIVRAKENAIIHAFIESLPNKYSTRVGDKGTQLSGGQKQRIALAPALVROPHEILLDE
              AYSALDSEKVVQALDQALRREGTCIVIAHRLSTIQNLADLIIVFQNGRVKHEGHTHQ
              LIAQKGIYFSMVSVOAGAKRQ"
BASE COUNT  1230 a 801 c 1039 g 1125 t
ORIGIN

Alignment Scores:
Pred. No.:      0      Length:      4195
Score:          5859.50      Matches:      1165
Percent Similarity: 94.80%      Conservative: 57
Best Local Similarity: 90.38%      Mismatches: 52
Query Match:      15      Indels: 15
DB:              6      Gaps: 5

US-09-672-725C-27 (1-1281) x AX108656 (1-4195)
Qy 1 MetAspProGluGlyArgLysGlySerAlaGlu---LysAsnPheTrpLysMetGly 19
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 ATGGATCTTGAAGGGGCCCAATGGAGGAGCAGAGAAGAAGAACTTTTAACTGAAC 159
Qy 20 LysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 160 AATAAAGT---AAAAAGATAAGAAAGAAAGAAACCACTGCTCAGTGTATTTCAATG 216
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 217 TTTCGCTATTCAAAATGGCTTGACAAGTGTATATGGTGGGGAACACTTTGGCGTCGCA 276
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 277 ATCCATGGAGCTGGACTTCTCTCATGATGCTGTGTGTTGGAGACATGACGGATACCTTT 336
Qy 80 AlaAsnAlaGly-----IleSerArgAsnLysThrPheProVal 92
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 337 GCAAATGCAGGAATAATTTAGGAGACTCTGTGTTGTTTAAACAACACT----- 387
Qy 93 IleIleAsnGluSerIleThrAsnAsnThrGlnHisPheIleAsnHisLeuGluGlu 112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 -----AATAGCAGTAATATCAGTATACATGATGCCCGTCATGAAT---CTGGAGGAAGAT 438
```

```
Qy 113 MetThrThrTyrAlaTyrTyrTyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyr 132
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439 ATGACCAGGTATGCTATTATTACAGTGAATTTGGTCTGGGTGCTGTTGCTGCTTAC 498
Qy 133 IleGlnValSerPheTrpCysLeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGln 152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 499 ATTCAAGTTTCATTTTGGTCTGGCAGCTGGAAGACAAATACACAAAAATAGAAAAACAG 558
Qy 153 PhePheHisAlaIleMetArgGlnGluIleGlyTrpPheAspValHisAspValGlyGlu 172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 559 TTTTTCATGCTATAATGCCAGACAGATAGGCTGGTTTGTATGTGCACGATGTTGGGGAG 618
Qy 173 LeuAsnThrArgLeuThrAspAspValSerLysIleAsnGluGlyIleGlyAspLysIle 192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 619 CTTAACACCCGGCTTACAGATGATGTCTCCAAGATTAATGAAGGAATTTGGTGCACAAATT 678
Qy 193 GlyMetPhePheGlnSerIleAlaThrPhePheThrGlyPheIleValGlyPheThrArg 212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 679 GGAATGTTCTTTTCTAGTCAATGGCAACATTTTTCCTACTGGGTTTATAGTAGGATTTTACACGT 738
Qy 213 GlyTrpLysLeuThrLeuValIleLeuAlaIleSerProValLeuGlyLeuSerAlaAla 232
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 739 GGTGGAAGCTTAACCTTGTGATTTTGCCCATCAGTCTCTTCTTGGACTGTCAAGTGA 798
Qy 233 IleTrpAlaLysIleLeuSerSerPheThrAspLysGluLeuLeuAlaTyrAlaLysAla 252
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 799 GTCTGGGCAAGATACTGCTTCTCATTTACTGATAAAGAACTCTTAGCTTATGCAAAAAGCT 858
Qy 253 GlyAlaValAlaGluGluValLeuAlaAlaIleArgThrValIleAlaPheGlyGlyGln 272
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 859 GGAGCAGTAGCTGAAGAGGCTCTGGCAATTAAGAATGTGATTTGCTTGGAGGACAA 918
Qy 273 LysLysGluLeuGluArgTyrAsnLysAsnLeuGluAlaLysGlyIleGlyIleLys 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 919 AAGRAAGAACTCGAAGGTACACAAAAATTTAGAAGAAGCTAAAGAATTTGGGATAAG 978
Qy 293 LysAlaIleThrAlaAsnIleSerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyr 312
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 979 AAAGCTATTACAGCCAATATTCTATAGGTGCTGCTTCTCTTATCTATGCAATCTTAT 1038
Qy 313 AlaLeuAlaPheTrpTyrGlyThrSerLeuValLeuSerSerGluTyrThrIleGlyGln 332
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1039 GCTCTGGCTCTCTGCTATGGGACCACCTTGGCTCTCAAGGAATATCTTATTGGACAA 1098
Qy 333 ValLeuThrValPhePheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerPro 352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1099 GTACTCACTGTATCTTTCTGTATTAATTGGGGCTTTTAGTGTGGACAGGCATCTCCA 1158
Qy 353 SerIleGluAlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheLysIleIleAsp 372
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1159 AGCATTTGAAGCATTTGCAAAATGCAAGAGGAGCAGCTTTTGAATCTTCAAGATTAATGAT 1218
Qy 373 AsnLysProSerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGly 392
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1219 AATAAGCAAGTATTGACAGCTATTTCGAAGAGTGGGCACAAACCAGATAATATTAAAGGA 1278
Qy 393 AsnLeuGluPheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeu 412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1279 AATTTTGAATTCAGAAATGTTTCACTTACCTTACCCTATCCCAATCGAAAAAGAAAGTTTAAAGATCTTG 1338
Qy 413 LysGlyLeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGly 432
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1339 AAGGCCCTGAACCTCAAGGTGCAGAGTGGGCGAGCGGTGGCCCTGGTGTGGAACAGCGGC 1398
Qy 433 CysGlyLysSerThrThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMet 452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1399 TGTGGAGAGGCACAAACCGTCCAGCTGATGACAGAGCTTTTATGACCCACACAGAGGCGCATG 1458
Qy 453 ValCysIleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThr 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1459 GTCAGTGTGTGGACAGGATATTAGGACCATAAACGTAAGGTTTCTACGGGAATATCATC 1518
```

QY 473 GlyValValSerGlnLeuProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyr 492
|||||
Db 1519 GGTGGTGGTGCAGTCAAGAACTGTATTGTTGGCCACCATAGAGTGAAGAACTTCGCTAT 1578

QY 493 GlyArgGluAsnValThrMetAspGluIleGluLysAlaValLysGluAlaAsnAlaTyr 512
|||||
Db 1579 GGTCTGAAGATGTCACCATGGATGAGATTGAGAAAGTGTCAAGAAAGCCAAATGCCCTAT 1638

QY 513 AspPheIleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGln 532
|||||
Db 1639 GACTTTATCATGAACCTGCCTCAGAAATTTGACACCCCTGGTTGGAGAGAGGGCCCCAG 1698

QY 533 LeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLys 552
|||||
Db 1699 CTGAGTGGTGGCAGCAAGCAGAGGATGCCATTGACCGTGGCTTGGTTCGCAACCCCAAG 1758

QY 553 IleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGln 572
|||||
Db 1759 ATCCCTCCCTGGTGGAGCGGCCAGCTCAGCCCTGGACACAGAAAGTGAAGCAGTGGTTCCAG 1818

QY 573 ValAlaLeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSer 592
|||||
Db 1819 GTGGCTCTGGATAGGCCAGAAAGTCCGACCATTTGTAGTACCTCATCGTTGTCT 1878

QY 593 ThrValArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGluLysGly 612
|||||
Db 1879 ACGGTTCGTAATCCGACGTCATCGCTGGTTTCGATGATGAGTCAITGTGTGGAAAGGA 1938

QY 613 AsnHisAspGluLeuMetLysGluLysGlyIleThrPheLysLeuValThrMetGlnThr 632
|||||
Db 1939 AATCATGATGAGCTCATGAAAGAGAAAGGCATTTACTTCAAACTTGTCACAATGCAGACA 1998

QY 633 ArgGlyAsnGluIleGluLeuGluAsnAlaThrGlyGluSerLysSerGluSerAspAla 652
|||||
Db 1999 CGAGAAATGAATGAATGAATGAATGACGCTGATGAATCCAAAGTGAATTTGATACC 2058

QY 653 LeuGluMetSerProLysAspSerGlySerSerLeuIleLysArgArgSerThrArgArg 672
|||||
Db 2059 TTGGAATGTCTTCATGATATTCAGGATCCAGTCTAATAAGAAAGATCCACTCGTAGG 2118

QY 673 SerIleHisAlaProGlnGlyAlaAspArgLysLeuGlyThrLysGluAspLeuAsnGlu 692
|||||
Db 2119 AGTGTCCGTGGATCACAAGGCCAAGACAGAAAGCTTTAGTACCAAGAGGCTCTGGATGAA 2178

QY 693 AsnValProValSerPheThrArgIleLeuLysLeuAsnSerThrGluThrProTyr 712
|||||
Db 2179 AGTATACCTCCAGTTCTCTTTGGAGGATTAATGAAGCTAAATTAATGAGTGGCCCTAT 2238

QY 713 PheValValGlyIlePheCysAlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIle 732
|||||
Db 2239 TTTGTGTGTGTGTATTGTGGCCATTATAATGGAGGTCTGCAACAGCATTTGTCAGTA 2298

QY 733 IlePheSerArgIleIleGlyIlePheThrArgAspGluAspProGluThrLysArgGln 752
|||||
Db 2299 ATATTTCAAAGATATAGGGATTTTACAGAAATATGATGCGCAACAAACACGACAG 2358

QY 753 AsnSerAsnMetPheSerValPheLeuValLeuGlyIleIleSerPheIleThrPhe 772
|||||
Db 2359 AATAGTAACCTGTTTTCATATTGTTCTAGTCTCTGGAATGTGTTCTTTTATTACATTT 2418

QY 773 PheLeuGlnGlyPheThrPheGlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyr 792
|||||
Db 2419 TTCCTTCAGGCTTCACATTTGGCAAGCTGGAGAGATCCTCACCAAGCGCTCCGATAC 2478

QY 793 MetValPheArgSerMetLeuArgGlnAspValSerTrpPheAspAspProLysAsnThr 812
|||||
Db 2479 ATGTTTTCCGATCCATGCTCAGACAGATGTGAGCTGGTTGTATGACCCCTAAACACAC 2538

QY 813 ThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGly 832
|||||
Db 2539 ACTGGACATGTACTACAGGCTCGCCAATGATGCTCTCAAGTTAAAGGGGCTATAGGT 2598

QY 833 SerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSer 852
|||||

Db 2599 TCCAGGCTGCTATAAATACCAGAAATATAGCAAAATCTTGGACAGGAATAAATATATCC 2658

QY 853 LeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAlaIleValProIleIleAlaIle 872
|||||
Db 2659 TTAATCTATGTTGGCAACTGACACTGTTACTCTTAGCAATTTGACCATCATTTGCAATA 2718

QY 873 AlaGlyValValGluMetLysMetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeu 892
|||||
Db 2719 GCAGAGTGTGTTGAATGAAATGTTGCTGGACACAGCACTGAAAGATAAGAAAGAACTA 2778

QY 893 GluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeu 912
|||||
Db 2779 GAAGTCTCTGGAGAGATCGCTACTGAAGCAATAGAAACTTCGAACTGTTGTTCTTCTTG 2838

QY 913 ThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsn 932
|||||
Db 2839 ACTCAGGAGCAGAAGTTTGAACATATATGATCATCAGAGTTTGCAGGTACCATACAGAAAC 2898

QY 933 SerLeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyr 952
|||||
Db 2899 TCITTTGAGGAAGACACATCTTTTGAATCAGCTTTTCCCTTCAGCGGCAATGATGTAT 2958

QY 953 PheSerTyrAlaGlyCysPheArgPheGlyValTyrLeuValAlaAsnGluPheMetAsn 972
|||||
Db 2959 TTTTCTCTATCTGATGTTTCCGGTTTGGAGCCTACTTGGTGGCACAATAGTCTCATGAGC 3018

QY 973 PheGlnAspValLeuLeuValPheSerAlaIleValIlePheGlyAlaMetAlaValGln 992
|||||
Db 3019 TTTGAGGATCTCTGTTAGTATTTTTCAGCTGTTGCTTTGTTGGTGCATGCGCGTGGGCAA 3078

QY 993 ValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMet 1012
|||||
Db 3079 GTCAGTTTCATTTGCTCTGACTATGCCAAAGCAAAAGTAGTATCAGACGCCACATCATG 3138

QY 1013 IleIleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThr 1032
|||||
Db 3139 ATCATTTGAAAAACCCCTTTGATGACAGCTACAGCACAGAGCCCTAAAGCCGACACA 3198

QY 1033 LeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIle 1052
|||||
Db 3199 TTGGAAGGAAATTCATATTAATGAAGTTGTTATCAACTATATCCACCGACTGGACATC 3258

QY 1053 ProValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrLeuAlaLeuValGly 1072
|||||
Db 3259 CCAGTCTTCAGGGCTGAGCCTGGAAGTGAAGAGGCGCAGCGCTGCCCTGGTGGGC 3318

QY 1073 SerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeu 1092
|||||
Db 3319 AGCAGTGGCTGTGGGAAGAGCAGCGGTGGTCCAGCTCTCTGGAGCGGTCTATGAGCCCTTG 3378

QY 1093 AlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArg 1112
|||||
Db 3379 GCGGGAAAGTGTGCTTGACGCCAAAGAAATAAAGCAACTGAATGTTTCAGTGGCTCCGA 3438

QY 1113 AlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsn 1132
|||||
Db 3439 GCACACTGGGCATCTGTCTCCAGGAGCCCATCTCTGTTTACTGTCAGCATTTAGTGAGAAC 3498

QY 1133 IleAlaTyrGlyAspAsnSerArgValValSerHisGluIleValGlnAlaAlaLys 1152
|||||
Db 3499 ATTGCCCTATGGAGACAACAGCGGGTGGTGTACAGGAAGAGATCGTGAGGCGCACCCAG 3558

QY 1153 GluAlaAsnIleHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGly 1172
|||||
Db 3559 GAGGCCAAATATACACGCCCTTCATCGAGTCACTGCCCTAATAAATATAGCACCAGATGAGA 3618

QY 1173 AspLysGlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeu 1192
|||||
Db 3619 GACAAAGGAACCTCAGCTCTCTGGTGCCAGAAACACGCAATTCATGATGCTGTCGCCCTT 3678

QY 1193 ValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer 1212
|||||

QY	480	ValLeupPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499
DB	1432	GTATTTGTTGGCCACCACGATAGCTGAAACCATTCGCTATGCCCTGAAATGTCCACCATG	1491
QY	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro	519
DB	1492	GATGAGATTGAGAAAGCTGTCAAGGAAGCCAAATGCCCTATGACTTTATCATGAAATGCCT	1551
QY	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539
DB	1552	CATAAATTGCACCCCTGTTGGAGAGAGGGGCCAGCTGTGAGTGGTGGGCAGAAAGCAG	1611
QY	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAla	559
DB	1612	AGGATGCCCATTCGACGTCGCCCTGGTTGCGAACCCCAAGATCCTCTGCTGGATGAGGCC	1671
QY	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579
DB	1672	ACGTGACGCTTGGACACAGAAAGCAGCAGTGGTTTCAGTGGCTCTGGATAAGGCCAGA	1731
QY	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599
DB	1732	AAAGTCGCGACCCACCATTTGATAGCTCATCGTTGCTGTACAGTTCGTAATGCTGACGTC	1791
QY	600	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619
DB	1792	ATCGCTGTTTCGATGATGAGATCATTTGGGAAAGGAATAATCATGATGAATCATGAAA	1851
QY	620	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu	639
DB	1852	GAGAAAGCATTTACTTCAAACCTGTGAGAAAGGAATAATCATGATGAATGAATTA	1911
QY	640	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	659
DB	1912	GAAATGCAGCTGATGAATCCAAAGTGAATGATGATGCTTGGAAATGCTCTCAAATGAT	1971
QY	660	SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly	679
DB	1972	TCAAGATCCAGCTTAATAAGAAAAAGATCAACTCGTAGGAGTGCCTGGATCACAAGCC	2031
QY	680	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe	699
DB	2032	CAAGACAGAAGCTTAGTACCAAGAGGCTCTGGAATGAAGATATACCTCCAGATTCTCTT	2091
QY	700	TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	719
DB	2092	TGGAGGATTATGAAGCTAAATTAACTGATGGCCTATTGTTGTTGGTGATTTTGT	2151
QY	720	AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	739
DB	2152	GCCATTATAATGGAGGCTCGAACACGATTTGCAATAATATTTTCAAAGATTATAGG	2211
QY	740	IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal	759
DB	2212	GTUTTTTACAGAATTGATGATCCTCGAACAAACACACAGATAGTAACTTGTTTTCACTA	2271
QY	760	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	779
DB	2272	TTGTTTCTAGCCCTTGGAAATATTTCTTTTATACATTTTTCCTTCAAGGTTTTCACATTT	2331
QY	780	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgThrMetValPheArgSerMetLeu	799
DB	2332	GGCAAGCTGGAGAGATCCTCACCAAGCGGCTCCGATACATGGTTTTCGATCCATGCTC	2391
QY	800	ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg	819
DB	2392	AGACAGGATGTGAGTTGGTTGATGATCCCTAAAAACACCACCTGAGCATTTGACTACCAG	2451
QY	820	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	839
DB	2452	CTCGCCAAATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTGCTGTAATACC	2511

Qy	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleValGlyTrpGlnLeu	859
Db	2512	CAGAAATATACAAACTCTGGGACAGGAATAATTATATCCTTCATCTATGGTTGGCAACTA	2571
Qy	860	ThrLeuLeuLeuAlaTleValProIleIleAlaIleAlaGlyValValGluMetLys	879
Db	2572	ACACTGTACTCTTAGCAATGTACCACATANTGCATAGCAGAGTGTCTTGAAATGAAA	2631
Qy	880	MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla	899
Db	2632	ATGTTGCTCTGGACAAGCACTGAAAGATAAGAAGAACTAGAGGTGCTGGGAAGATCGCT	2691
Qy	900	ThrGluAlaIleGluAsnPheArgThrValValSerIleuThrArgGluGlnLysPheGlu	919
Db	2692	ACTGAAGCAATAGAAAACTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAAGTTTGAA	2751
Qy	920	TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle	939
Db	2752	: : : : : CATATGATGCTCAGAGTTTCAGGTACCATACAGAACTCTTTGAGGAAGCACACATC	2811
Qy	940	PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe	959
Db	2812	: : : : : TTTGGAAATTACATTTTCCTCACCCAGGCCAATGATATATTTTTCTATGCTGGATGTTTC	2871
Qy	960	ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal	979
Db	2872	: : : : : CGSTTTGGAGCCTACTGGTGGCACATAAACACTCATGAGCTTTGAGGATGTTCTGTTAGTA	2931
Qy	980	PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp	999
Db	2932	TTTTTCAGCTGTGTCTTTGTTGGTCATGGCGCTGGGCGAAGTCAGTTTCATTTGCTCCTGAC	2991
Qy	1000	TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu	1019
Db	2992	: : : : : TATGCCAAGCCAAAATATCAGCAGCCCACATCATCATGATCATTTGAAAAAACCCCTTTG	3051
Qy	1020	IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe	1039
Db	3052	AFTGACAGCTACAGCAGGAAGCGCTTAATGCCGCAACACATTTGGAAGAAATGTACATTT	3111
Qy	1040	AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer	1059
Db	3112	GCTGAAGTTGATTCAACTATCCCCACCCCGGACATCCCGAGTGCCTCAGGAGCACTGAGC	3171
Qy	1060	LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer	1079
Db	3172	CTGGAGGTGAAGAAGGCCACAGCGCTGGCTCTGTTGGCGCACGCTGGTGGGAAGAGC	3231
Qy	1080	ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp	1099
Db	3232	: : : : : ACAGTGTGCAGCTCTCTGGAGGGTGTCTACGACCCCTTGGCAGGGAAGTGTCTGTGAT	3291
Qy	1100	GlyLysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSer	1119
Db	3292	GGCAAAGAAATAAAGCGACTGAATGTTCAGTGGCTCCGAGCACACCTGGGCATCGTGCC	3351
Qy	1120	GlnGluProIleLeuPheAspcysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer	1139
Db	3352	CAGAGGCCCATCTGTTTACTGCAGCATTCGCTGAGAACATTCGCTATGAGACAACAGC	3411
Qy	1140	ArgValValSerHisGluGlutIleValGlnAlaLysGluAlaAsnIleHisHisPhe	1159
Db	3412	: : : : : CGGTTGGTGTACAGGACAGAGATCGTGAGGCGACGAAGAGGCCACATACATGCGCTTC	3471
Qy	1160	IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer	1179
Db	3472	: : : : : ATCGAGTCACTGCCTTAATAATATAGCACTAAAGTAGGAGCAACAAGAACTCAGCTCTCT	3531
Qy	1180	GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu	1199
Db	3532	: : : : : GGTGGCCAGNAACAGGCATTCGCATAGCTCGTCCCTTGTAGCAGGCTCATATTTTG	3591
Qy	1200	LeuLeuAspGluAlaThrserAlaLeuAspThrGluSerGluLysValValGlnGluAla	1219


```
|||||
Db 3592 CTTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTGTCTCCAGAGCC 3651
|||||
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 3652 CTGGCAAAAGCCAGAGAGCCGACCTGCATGTGATTGCTCACCCTGTCCACCATC 3711
|||||
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||
Db 3712 CACAATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGAGTCAAGGAGCATGGCAGCAT 3771
|||||
Qy 1260 GlnGlnLeuAlaGlnLysGlyIleThrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 3772 CAGCAGCTGTCGACAGAAAGGCATCTATTTTTCATGGTCTCAGTGTCCAGGCTGGAACA 3831
|||||
Qy 1280 LysArg 1281
|||||
Db 3832 AAGCCG 3837
|||||
RESULT 14
AX336420 4646 bp DNA linear PAT 09-JAN-2002
LOCUS AX336420
DEFINITION Sequence 6929 from Patent WO0194629.
ACCESSION AX336420
VERSION AX336420.1 GI:18127139
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Horrigan,S., Soppet,D.R. and Weaver,Z.
Young,P.B., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 6929 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1. 4646
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1371 a 892 c 1129 g 1254 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4646
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 6 Gaps: 4
US-09-672-725C-27 (1-1281) x AX336420 (1-4646)
Qy 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
|||||
Db 425 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAGAGAACTTTTAAACTGAAC 484
|||||
Qy 20 LysLysSerLysLysLysGluLysLysLysProThrValSerThrPheAlaMet 39
|||||
Db 485 AATAAAGT---GAAAGATAGAAGGAAGAAAGAAACCACTGTCAGTGTATTTCAATG 541
|||||
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
|||||
Db 542 TTTTCGCTATTCAANTGGCTTCACAAGTTGTATATGGTGGTGGGAACCTTGGCTGCCATC 601
|||||
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
|||||
Db 602 ATCCATGGGGCTGGAGCTTCTCTCATGATGCTGGTGTGGAGAAATGACAGATCTCTT 661
|||||
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
|||||
Db 662 GCNAATGCAGGA---AATTAGAGATCTGTATGTCACCAATCACTAATAAGAGTGTATATC 718
```

```
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrThrAlaTyrTyr 119
|||||
Db 719 AATGATACAGGGTTCTTCATGAAT---CTGAGGAACACATGACCAGGTATGCTATATAT 775
|||||
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
|||||
Db 776 TACAGTGAATTTGGTGTGGGGTGTCTGCTTACATTTCAGGTTTTCATTTTGGTGC 835
|||||
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
|||||
Db 836 CTGGCAGCTGGAAGCAAAATACAAAATTAGAAAACAGTTTTTTCATGCTATATAATCGGA 895
|||||
Qy 160 GlnGluIleGlyTyrPheAspValHisAspValGlyGluLeuLeuAsnThrArgLeuThrAsp 179
|||||
Db 896 CAGGAGATAGGCTGCTTTGATGTGCACGATCTTGGGAGCTTAACACCCGACTTTACAGAT 955
|||||
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
|||||
Db 956 GATGTCTCTAAGATTAAATGAAGTTATTGGTGACAAAATTTGGAATGTTCTTTCAGTCAATG 1015
|||||
Qy 200 AlaThrPheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
|||||
Db 1016 GCACATTTTTTTCACGTGGGTTTATAGTAGGATTACACGCTGTTGGAAGCTAACCCCTGTG 1075
|||||
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
|||||
Db 1076 ATTTTGGCCATCAGTCTCTTCTTGGACTGTGCTGGGCAAGATACATCTCT 1135
|||||
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
|||||
Db 1136 TCATTTACTGTAAAGAACTCTTACGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTGC 1195
|||||
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
|||||
Db 1196 TTGGCAGCAATTAGAACTGTGATTGCAATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 1255
|||||
Qy 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
|||||
Db 1256 AACAAAATTTAGAAAGCTAAAAGAAATGGGATAAAGAAAGCTATTACAGCCCAATATT 1315
|||||
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
|||||
Db 1316 TCTATAGGTGCTGCTTCTGCTGATCTATGATCTTATGCTCTGGCTTCTGGTATGGG 1375
|||||
Qy 320 ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer 339
|||||
Db 1376 ACCACCTTGGTCTCTCAGGGGAATATCTATTTGGACAAAGTACTCACTGTATCTTTCT 1435
|||||
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
|||||
Db 1436 GTATTAAATTTGGGCTTTTAGTGTGGACAGGCATCTCCAAGCATTTGAAGCATTTGCAAAAT 1495
|||||
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
|||||
Db 1496 GCAAGAGGAGCAGCTATGAATCTTCAAGATAAATTTGATAAATGAAGCCAAAGTATTGACAGC 1555
|||||
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
|||||
Db 1556 TATTGGAAGAGTGGGCAAAACCAACAGATAATTTAAAGGAAATTTGGAATTCAGAAATGTT 1615
|||||
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
|||||
Db 1616 CACTTCAGTTACCCATCTCGAAAAGAAAGTTAAGATCTTGAAGGGCTGAACCTGAAGGTG 1675
|||||
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
|||||
Db 1676 CAGAGTGGGAGAGCGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAGAGACACACAGTCT 1735
|||||
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
|||||
Db 1736 CAGCTGATGCAGAGGCGCTATATGACCCACAGAGGGGATGCTGAGTGTGTGATGCAGGAT 1795
```


QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
DB 1796 ATTAGACCATAAATGTAAGGTTTCTACGGAAATCAITGGTGGTGGTCAAGAACT 1855
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
DB 1856 GTATTGTTGGCCACCACGATAGCTGAAACCATTCGCTATGGCGGTGAAATGTCACCATG 1915
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
DB 1916 GATGAGATTGAGAAGCTGTCAAGAGCCAAATGCCTATGACTTATCATGAAATGCGCT 1975
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
DB 1976 CATAAATTGACACCTGGTGGAGAGAGGGGCCAGTTGAGTGGTGGCAGAGGAG 2035
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
DB 2036 AGGATCGCCATTGCGACGTGCCCTGTTGCGAACCCCAAGATCCTCCTGCTGGATGAGGCC 2095
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
DB 2096 AGCTCAGCCTTGCACACAGAACGACAGTGGTTCAGGTGGCTCTGGATAAGGCCAGA 2155
QY 580 LysGlyArgThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
DB 2156 AAAGTCGCGACCCATTGTGATAGCTCATCGTTGTCTACAGTTCTGTAATGCTGACGTC 2215
QY 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
DB 2216 ATFCGCTGGTTTCGATGATGGATCTATTGTTGAGAAAGGAAATCATGATGAATCATGAA 2275
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
DB 2276 GAGAAGGCATTACTTCAACTTGTCAATGTCACATGACAGCAGGAGAAATGAAGTTGAATTA 2335
QY 640 GluAsnAlaThrGlyLysSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
DB 2336 GAAATGCGAGCTGATGATCCAAAGTCAAATTTGATGCTTGGAAATGCTTCAAATGAT 2395
QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
DB 2396 TCAAGATCCAGTCTAATAAGAAAGATCAACTCGTAGAGTGTCCGTGGATCAACAGCC 2455
QY 680 GlnAspArgLysLeuGlyLysGluAspLysGluAsnGluAsnValProProValSerPhe 699
DB 2456 CAAGACACAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCCTT 2515
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
DB 2516 TGGAGGATTATGAAGCTAAATTAATCAATGGCCTTATTTTGTGTTGGTGTATTGT 2575
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
DB 2576 GCCATTAAATGGAGGCTGCAACAGCATTTCGAATAATATTTCAAAGATTATAGGG 2635
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
DB 2636 GTTTTACAAAGATTGATGATCCCTGAAACAAACGACAGATAGTAATCTTTTCTACTA 2695
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
DB 2696 TTGTTTCTAGCCCTTGGAAATTATTTCTTTTATACATTTTTCCTTCAGGGTTTTCACATT 2755
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
DB 2756 GGCAAGCTGGAGAGATCCTCACAAGGGCTCCGATACATGGTTTCCGATCATGCTC 2815
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
DB 2816 AGACAGGATGTGAGTGTGTTGATGACCTTAAACACCACTGGAGCATTTGACTACCAAG 2875
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839

DB 2876 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTGTGTAATTACC 2935
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleIleIleIleIleIle 859
DB 2936 CAGAAATAGCAAAATCTTGGACAGGAATAATTATATCTTCTATCTATGTTGGCAACTA 2995
QY 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
DB 2996 ACCTGTTTACTTCTTAGCAATTGTACCCCATCATTGCAATAGCAGGATGTTGAAATGAAA 3055
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
DB 3056 ATGTTGCTGGACAAGCACGTGAAGATAAGAAAGAACTAGAGGTGCTGGGAAGATCGCT 3115
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
DB 3116 ACTGAAGCAATAGAAAACCTCCGAACCGTTGTTTCTTGTGACTCAGGAGCAGAAATTTGAA 3175
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
DB 3176 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAGACACATC 3235
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
DB 3236 TTTGGAATTACATTTTCTTCCACCAGCAATGATGATATTTTCTTATGCTGGATGTTTC 3295
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
DB 3296 CGGTTTGGAGCTACTTGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 3355
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
DB 3356 TTTTCACTGTTGCTTGTGGTCCATGGCGTGGGCAAGTCACTTATTTGCTCTCTGAC 3415
QY 1000 TyrAlaLysAlaLysValSerAlaHisValIleMetIleIleGluLysSerProLeu 1019
DB 3416 TATGCCAAAGCCAAATATCAGCAGCCACATCATCATGATCATTTGAAATAACCCCTTG 3475
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
DB 3476 ATTGACAGCTACACGCGAAGGCCCTAATGCCGAACACATTTGGAAGAAATGTCACATT 3535
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
DB 3536 GGTAAGTTGATTCACTATCCACCACCGACATCCAGTGTCTTCAGGAGCTGAGC 3595
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
DB 3596 CTGGAGGTGAAGAAGGCCACAGACGCTGGCTCTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3655
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
DB 3656 ACATGTGCTCAGCTCTCGAGCGGTCTACAGCCCTTGGCAGGAAAGTGTCTGTGTAT 3715
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
DB 3716 GGCAAGAAATAAGCACTGAAATGTTACGTGGCTCCGAGCACACCTGGGCATCGTGTCC 3775
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
DB 3776 CAGGAGCCATCTCTGTTGACTGCAGCATTTGCTGAGAACATTTGCCATTTGGAGACAACAGC 3835
QY 1140 ArgValValSerHisGluGluIleValGlnAlaLysGluAlaAsnIleHisIlePhe 1159
DB 3836 CGGTGGTGTACAGGAAGAGATCGTGAGGGCAGCAAGAGGACCAACATACATGCTCTC 3895
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
DB 3896 ATCGAGTCACTGCTTAATAATATAGCCTTAAAGTAGGAGACAAAGAACTCAGCTCTCT 3955
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199

```
Db 3956 GGTGGCAGAAACACAGCATTTGCCATGCTGCTGCCCTTGTAGACAGCCTCATATTTTG 4015
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAAGGTTCTCCAAAGGCC 4075
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGACAAAGCCAGAGAGCCGACCTGCATGTGATTCCTCAGCCCTGTCACCATC 4135
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGAATGCAGACTTAATAGTGTGTTTTCAGATGGCAGAGTCAAGGAGCATGGCAGCAT 4195
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleThrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGTCGACAGAAAGGCATCTATTTTCATGCTGAGTGTCCAGGCTGGAACA 4255
Qy 1280 LysArg 1281
Db 4256 AAGCGC 4261

RESULT 15
AX336708
LOCUS AX336708 4646 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 7217 from Patent WO0194629.
ACCESSION AX336708
VERSION AX336708.1 GI:18127427
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrikan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 7217 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
location/Qualifiers
1..4646
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1371 a 892 c 1129 g 1254 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 4646
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 6 Gaps: 4

US-09-672-725C-27 (1-1281) x AX336708 (1-4646)
Qy 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 425 ATGATCTTGAAGGGACCCCAATGAGGAGGCAAGAAAGAAAGCACTTTTAACTGAC 484
Qy 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 485 AATAAAAGT--GAAAAAGATAAGAGGAAAGAAACCAACATGTCAGTGTATTTTCAATG 541
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
Db 542 TTTTGGCTATTCAAATTTGGCTTTGACAAAGTTGATATGGTGGGAACTTTGGCTGCCATC 601
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspPhe 79
Db 602 ATCCATGGGCTGGACTTCCTTCATGATGCTGGTGTGGAGAAATGACATATCTTT 661
```

```
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleLeuAsnGluSerIleThr 99
Db 662 GCAAATGCAGCA---AATTAGAAAGATCTGATGTCAAACATCACTAATAAGAGTGATATC 718
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 719 AATGATACAGGGTTCTTCATGAAT--CTGGAGGAAGACATGACCAGGTATGCGCTATAT 775
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaIleValIleGlnValSerPheTrpCys 139
Db 776 TACAGTGAATTTGGTCTGGGCTGCTGCTGCTTACATTCAGGTTTCATTTTGGTGC 835
Qy 140 LeuAlaIleGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 836 CTGGCAGCTGGAAGACAAATACAAAAATTAGAAAAACAGTTTTTTTCATGCTATATAATCG 895
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 896 CAGGAGATAGGCTGGTTGATGTGCACGATTTGGGAGCTTTAACACCCGACTTACAGAT 955
Qy 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
Db 956 GATGCTCTAAGATTAAAGATTATTGGTCACAAATTTGGAATGTTCTTTCAGTCAATG 1015
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 1016 GCAACATTTTTCACCTGGGTTTATAGTAGGATTTACACGTGTTGGAAAGCTAACCCCTTGTG 1075
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 1076 ATTTTGGCCATCAGTCCCTGTTCTTGGACTGTCAGCTGCTGCTGGCAAGATACTATCT 1135
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal 259
Db 1136 TCATTACTGATAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTGC 1195
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 1196 TTGGCAGCAATTAGAACTGTGATTGCAATTTGGAGGCAAAAAGAAAGAACTTTGAAGGTAC 1255
Qy 280 AsnLysAsnLeuGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle 299
Db 1256 AACAAAAATTAGAAAGAACTAAAAGAAATTTGGGATAAAGAAAGCTNTTACAGCCATATT 1315
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrTyrGly 319
Db 1316 TCTATAGTGTGCTTCTCTGCTGATCTATGCATCTTATGCTCTGGCTTCTGTTATGGG 1375
Qy 320 ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer 339
Db 1376 ACCACCTTGGTCTCTCAGGGGAATATTCTATTGGACAAGTACTCAGCTGATATTCTTTCT 1435
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1436 GTATTAAATTTGGGCTTTTAGTGTGGACAGCATCTCCACCATTTGAAGCATTTGCAAT 1495
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleLeuAspAsnLysProSerIleAspSer 379
Db 1496 GCAAGAGCAGCAGCTTATGAAATCTTCAAGATAATTGATAATAAGCCAAAGTATTGACAGC 1555
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1556 TATTTCGAAGAGTGGGCACAAACCAAGATAATATTAAAGGAAATTTGGAATTTGAGAAATGTT 1615
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1616 CACTTCAGTTACCCATCTCGAAAAGAACTTAAGATCTTGAAGGGCCCTGAACCTGAAGGTG 1675
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1676 CAGAGTGGCAGACGCTGGCCCTGTTGGAAACAGTGGCTGTGGGAGAGACACACAGTCTC 1735
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
```

Db 1736 |||||CAGCTGATCAGAGGCTCTATACCCACAGAGGGGATGGTCAGTGTTCGATGACAGGAT 1795
QY 460 |||||IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1796 |||||ATTAGGACCATAAATGTAAGTTTCTACGGGAATCATTTGGTGGTGAAGTCCAGAACTT 1855
QY 480 |||||ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1856 |||||GTAATCTTTGCCACACGATAGCTGAAACATTCGCTATGCGCGTGAAATGTCACCATG 1915
QY 500 |||||AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1916 |||||GATGAGATTGAGAAAGCTCTCAAGGAAGCCAAATGCCTATGACTTATCATGAAACTGGCT 1975
QY 520 |||||AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnGln 539
Db 1976 |||||CATAAATTTGACACCTGTTGGAGAGAGAGGGGCCAGTTGAGTGGCGGAGAAGCAG 2035
QY 540 |||||ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 2036 |||||AGGATCGCCATTGCACGTGCCCTGGTTGCCAACCCCAAGATCCTCCTGCTGGATCAGGCC 2095
QY 560 |||||ThrSerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArg 579
Db 2096 |||||ACGTACGCCCTTGGACACAGAAAGCCAGAGCGTGTTCAGTGGCTCTGGATAAGGCCAGA 2155
QY 580 |||||LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 2156 |||||AAAGTCGAGCACCATTCATGTAAGTCTCATCGTTGTCTACAGTTCGTAAGTCTGACGTC 2215
QY 600 |||||IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 2216 |||||ATCGCTGGTTTCGATGATGGAGTCATTTGGAGAAAGGAATCATGTAAGTCAATGAAA 2275
QY 620 |||||GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 2276 |||||GAGAAGGCAATTTACTTCAAACTTCTCAATGTCAGACAGCAGCAAGAAATCAAGTTGAATA 2335
QY 640 |||||GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2336 |||||GAAATGCGAGCTGATGAATFCCAAAAGTGAATGTAGTGGCTTGGAAATGTCCTCAAAATGAT 2395
QY 660 |||||SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 2396 |||||TCAAGATCCAGTCTAATAAGAAAAAGATCACTCGCTAGAGTGTCCGTGGATCACAAGCC 2455
QY 680 |||||GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
Db 2456 |||||CAAGACAGAAAGCTTAGTACCAGAGGCTCTGGATGAAGTATACCTCCAGTTTCCCTTT 2515
QY 700 |||||TrpArgIleLeuLysLeuAsnSerThrGluThrProTyrPheValValGlyIlePheCys 719
Db 2516 |||||TGGAGATTATGAGACTAAATTAATACTGAATGGCCCTAATTTGTTGGTGTATTTTGT 2575
QY 720 |||||AlaIleLeuAsnGlyGluGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2576 |||||GCCATTATAAATGGAGGCTGCAACCCAGCATTTGGCAATAATATTTCAAGATATATAGG 2635
QY 740 |||||IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 |||||GTTTTTACAAGAAATTTGATGATCTCTGAAACAAAACAGACAGAAATAGTAATCTGTTTCTACTA 2695
QY 760 |||||LeuPheLeuValIleGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2696 |||||TTGTTTTCAGCCCTTGGAAATTAATTTCTTTTATATATTTTCTTCAGGGTTTCACATTT 2755
QY 780 |||||GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 |||||GGCAAGCTGGAGAGATCCTCACCAAGCGGCTCCGATACATGGTTTCCGATCCATGCTC 2815
QY 800 |||||ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819

Db 2816 |||||AGACAGGATGTGAGTGTGGTTTGTATGACCTAAAAACACCACCTGAGGACATTGACTTACCAGG 2875
QY 820 |||||LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2876 |||||CTCGCAATGATGCTGCTCAAGTTTAAAGGGCTATAGTTTCCAGGCTTGTCTGTAATTACC 2935
QY 840 |||||GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2936 |||||CAGAAATATAGCAAAATCTTGGGACAGGAATAATTATATCTTCACTCATGTTGGTGGCAACTA 2995
QY 860 |||||ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2996 |||||ACACTGTTACTCTTACCAATTTGTACCATCTATGCAATAGCAGGAGTTGTTGAAATGAAA 3055
QY 880 |||||MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 3056 |||||ATGTTGCTGCAAGCACTGAAAGATAAGAAAGAACTAGAAAGTGTGGGAAGATCGCT 3115
QY 900 |||||ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 3116 |||||ACTGAAGCAATAGAAAACTTCCGAACCGTGTCTTCTTGTACTCAGGACAGAAAGTTTGA 3175
QY 920 |||||TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 |||||CATATGTATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATC 3235
QY 940 |||||PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 |||||TTTGGAAATACATTTTCTTCCACCCAGCAATGATGATATTTTCTATGCTGGATGTTTC 3295
QY 960 |||||ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 3296 |||||CGGTTTGGAGCCTACTTGGTGGCACATAAACTCATGAGCTTGGAGGATGTTCTGTTAGTA 3355
QY 980 |||||PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3356 |||||TTTTCAGTGTGTCTTGGTGCCATGCGGTGGGCAAGTCACTTCAATTTGCTTCTCTGAC 3415
QY 1000 |||||TyrAlaLysAlaLysValSerAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 |||||TATGCCAAAGCCAAATATCAGCCACCCACATCATGATCATTTGAAAACCCCTTTG 3475
QY 1020 |||||IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 |||||ATTGACAGCTACACCCGGAAGGCTTAATGCCGAACACATTTGGAGGAATATGCACATTT 3535
QY 1040 |||||AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 |||||GGTCAAGTTGTATTCAACTATCCACCCGACCGACATCCAGTGTCTCAGGGACTGAGC 3595
QY 1060 |||||LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3596 |||||CTGGAGGTGAAGAGGCCAGACGCTGGCTGTGTGGGACAGAGTGGCTGTGGGAAGAGC 3655
QY 1080 |||||ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3656 |||||ACAGTGTGCCAGCTCCTGGAGCGGTTCTACGACCCCTTGGCAGGAAAGTGTCTGTAT 3715
QY 1100 |||||GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 |||||GGCAAGAAATAAAGCCACTGAATGTTCAGTGGCTCCGAGCACACCTGGGCATCGTCC 3775
QY 1120 |||||GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 |||||CAGAGGCCCATCTCTGTTGACTGCAGCATTTGCTGAGAACATTCCTATGGAGACACAGC 3835
QY 1140 |||||ArgValValSerHisGluGluIleValGlnAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3836 |||||CGGTGTGTGCACAGGAAGAGATCGTGAGGCGACGAAGAGGCCAACATACATGCTTC 3895
QY 1160 |||||IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 |||||ATCAGTCACTGCTTAATAAATATAGCACTAAAAGTAGGAGACAAAGAACTCAGCTCTCT 3955

QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
 Db 3956 GGTGGCCAGAAACACGCATTGCCATAGCTCGTCCCTTGTAGACAGCCTCATATTTG 4015
 QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
 Db 4016 CTTTGTGGATCAAGCCACGTCAGCTCTGGATACAGAAAGTGAAAAAGTTGTCCAAGAAGCC 4075
 QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
 Db 4076 CTGGACAAAGCCAGAGAAGCCGCACCTGCTGATTGCTCACCCTGTCCACCATC 4135
 QY 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
 Db 4136 CAGATGCACACTTAATAGTGGTTCAGATGGCAGAGTCAAGGAGCATGGCAGCAT 4195
 QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
 Db 4196 CAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 4255
 QY 1280 LysArg 1281
 Db 4256 AAGCGC 4261

Search completed: November 7, 2002, 05:09:40
 Job time : 4104.81 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:36:49 ; Search time 301.645 Seconds
(without alignments)
7291.243 Million cell updates/sec

Title: US-09-672-725C-27
Perfect score: 6473
Sequence: 1 MDPEGRKSAEKNFKMGK.....LLAQGIYFSMVSVQAGAKR 1281

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09672725/runat_04112002_124338_17877/app_query.fasta_1.6435
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=ring -MINMATCH=0 -1-LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09672725.ecgn_1.1.0 -runat_04112002_124338_17877 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	6473	100.0	4279	22	AAD03506
2	6466	99.9	4279	22	AAD03505
3	6461	99.8	4279	22	AAD03504
4	6456	99.7	4279	22	AAD03488
5	6425.5	99.3	4317	22	AAD03489
6	5866	90.6	3860	21	AZ49332
7	5866	90.6	3860	24	ABA94365
8	5862	90.6	4186	22	AAF86127
9	5859.5	90.5	4195	22	AAF86128
10	5857	90.5	3860	21	AZ49333
11	5857	90.5	3860	24	ABA94366
12	5857	90.5	4349	22	AAH57442
13	5857	90.5	4646	21	AZ94738
14	5857	90.5	6505	17	AAI13394
15	5857	90.5	8630	21	AZ24041
16	5857	90.5	8630	21	AZ24042
17	5854	90.4	4669	8	AAI70752
18	5851	90.4	4378	11	AAQ04522
19	5849.5	90.4	3988	21	AZ88973
20	5844	90.3	4646	15	AAQ27872
21	5840	90.2	4669	14	AAQ2726
22	5830	90.1	4669	19	AAV32645
23	5815.5	89.8	4264	19	AAV66533
24	5815.5	89.8	4264	19	AAV66534
25	5641	87.1	4788	21	ABA94335
26	5641	87.1	4788	24	ABA94368
27	5628	86.9	4425	21	AZ52048
28	5618	86.8	4369	21	AZ52047
29	5301	81.9	4189	24	ABA94367
30	5301	81.9	4189	24	ABA94367
31	5301	81.9	4313	14	AAQ38950
32	5291.5	81.7	4233	21	AAZ90198
33	5291.5	81.7	4233	22	AAZ7498
34	4922.5	76.0	3924	21	AAZ94742
35	4922.5	76.0	3924	21	AAZ94742
36	3631	56.1	2726	15	AAQ70907
37	3631	56.1	2726	15	AAQ70916
38	3631	56.1	2726	18	AAZ43322
39	3308.5	51.1	4776	21	AAZ94744
40	2824.5	43.6	4175	20	AAV69392
41	2695	41.6	3942	23	ABL14175
42	2561	39.6	4157	23	ABL06835
43	2514	38.8	4047	20	AAZ60201
44	2497	38.6	4111	23	ABL16205
45	2482	38.3	4002	20	AAZ85823

ALIGNMENTS

RESULT 1
AAD03506
ID AAD03506 standard; cDNA; 4279 BP.
XX
AC AAD03506;
XX
DT 13-JUN-2001 (first entry)
XX
DE Dog P-glycoprotein (PGP) allelic variant (Genotype D) cDNA.
XX
KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
KW MDRI; drug bioavailability; transgenic animal; genetic model; ss.
XX
OS Canis familiaris.
XX
FH Key
CDS 17..3862
/*tag= a

```

FT      /product= "Dog P-glycoprotein (PGP) allelic variant
FT      (Genotype D) protein"
FT      replace (31, T)
FT      /*tag= b
FT      allele
FT      allele
FT      replace (607, C)
FT      /*tag= c
FT      allele
FT      replace (1001, T)
FT      /*tag= c
FT      allele
FT      replace (3458, A)
FT      /*tag= c
FT      WO200123540-A2.
FT      PN
FT      XX
FT      XX
FT      05-APR-2001.
FT      PD
FT      XX
FT      XX
FT      28-SEP-2000; 2000WO-US26767.
FT      XX
FT      XX
FT      28-SEP-1999; 99US-0156510.
FT      PR
FT      XX
FT      PA
FT      (GENT-) GENTEST CORP.
FT      XX
FT      Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
FT      PI
FT      XX
FT      WPI: 2001-235373/24.
FT      DR
FT      P-PSDB; AAE00310.
FT      XX
FT      New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
FT      for determining the bioavailability of drugs and for screening for dog
FT      PGP inhibitors -
FT      XX
FT      PS
FT      Claim 9; Page 102-107; lilpp; English.
FT      XX
FT      The invention relates to dog P-glycoprotein (PGP) also referred
FT      as multidrug transporter (MDR1) and nucleic acids encoding them.
FT      CC
FT      The invention also includes fragments and biologically functional
FT      CC
FT      variants of dog P-glycoprotein. PGP and their nucleic acids are
FT      CC
FT      useful for determining the bioavailability of drugs and for
FT      CC
FT      screening PGP inhibitors. They are useful for the diagnosis and
FT      CC
FT      treatment of conditions characterised by PGP activity, by
FT      CC
FT      reducing or increasing PGP activity in a cell. PGP nucleic acids
FT      CC
FT      are used as oligonucleotide probes. Complements of PGP nucleic
FT      CC
FT      acids are useful as antisense oligonucleotides, to induce a PGP
FT      CC
FT      'knockout' phenotype. They are used to prepare a non-human
FT      CC
FT      transgenic animal, which are valuable as genetic models for
FT      CC
FT      human diseases.
FT      CC
FT      The present sequence is dog P-glycoprotein (PGP) allelic variant
FT      CC
FT      (Genotype D) cDNA. The PGP enzyme functions as an efflux pump
FT      CC
FT      exporting small molecules across the cell membrane. This enzyme
FT      CC
FT      is a member of the ABC transporter family.
FT      CC
FT      SQ
FT      Sequence 4279 BP; 1296 A; 833 C; 1009 G; 1141 T; 0 other;
FT      XX
FT      Alignment Scores:
FT      Pred. No.: 0 Length: 4279
FT      Score: 6473.00 Matches: 1281
FT      Percent Similarity: 100.00% Conservative: 0
FT      Best Local Similarity: 100.00% Mismatches: 0
FT      Query Match: 100.00% Indels: 0
FT      DB: 22 Gaps: 0
FT      XX
FT      US-09-672-725C-27 (1-1281) x RAD03506 (1-4279)
FT      QY
FT      1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
FT      Db
FT      17 ATGGATCTCTGAAGGAGCCGTAAAGGGAGTGCAGAGAGAAGACTCTCGAAAATGGGAAA 76
FT      QY
FT      21 LysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
FT      Db
FT      77 AAAAGTAAAAAAGAGAGAAGAAAAAGAAACCAACTGTGCAGCAGCTTTGCAATGTT 136
FT      QY
FT      41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleile 60
FT      Db
FT      137 CGCTATTCAAAATGGCTTGATAGTGTATATGTTGGTGGGACAAATGGCTGCATCATC 196

```

```

QY      61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
Db      197 CATGGAGCTGCACCTCCTCATGATGCTGGTTTTTGGAAACATGACAGATAGCTTTGCA 256
QY      81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db      257 AATGCAGGAATTTCAAGAAACAAACAACTTTCCAGTTATAATTAATGAAGTATATGGAAC 316
QY      101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 120
Db      317 AATACACACATTTTCATCAACCATCTGGAGGAGGAATGACCAGGTATGCTATTATTAC 376
QY      121 SerGlyIleGlyAlaGlyValLeuValAlaAlaIleIleGlnValSerPheTrpCysLeu 140
Db      377 AGTGGGATCGGTGCTGGCGTGTGGTGTGCTTACATCCAGGTTTCATTCTCGTGCCTG 436
QY      141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
Db      437 GCAGCAGGAAGACAGATACTCAAAATTTAGAAACAATTTTTCATGCTATCATCGCACAG 496
QY      161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
Db      497 GAGATTGGCTGGTTTCACGTGCATCGCTGGGGAGCTTAACACCCGGCTCAGACAGAT 556
QY      181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIleAla 200
Db      557 GTCTCCAAAATCAATGAAGGAATGGCGACAAATTTGGAATGTTCTTTCAATCAATAGCA 616
QY      201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
Db      617 ACATTTTTCACCCGTTTATAGTGGGGTTTACAGCTGGTGGGAAGCTTAACCTTGTGATT 676
QY      221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
Db      677 TTGGCCATCAGCCCTGCTTCTGGACTTTTCAGCCCATCTGGGCAAGAGTACTATCTTCA 736
QY      241 PheThrAspLysGluLeuLeuAlaIleValAlaLysAlaGlyAlaValAlaGluGluValLeu 260
Db      737 TTACTGATAAAGAACTCTTGCCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 796
QY      261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
Db      797 GCAGCAATCAGAACTGTGATTGCCCTTTGGAGGACAAAAGAAAGACTTGAAGGTACAAC 856
QY      281 LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
Db      857 AAAAATTAGAGAAGCTAAAGGAATTCGGATAAAGAAAGCTATCAGGCCCAACATTTCT 916
QY      301 IleGlyAlaAlaPheLeuLeuIleThrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
Db      917 ATTTGGTCCCGCTTTCTATTGATCTATCATCATATGCTCTGGCTTCTGGTATGGGACC 976
QY      321 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 340
Db      977 TCCTTTGCTCTCTCCAGTGAATATATCTATTGGACAAGTACTCAGCTGCTCTCTTTCTGTA 1036
QY      341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db      1037 TTAATTGGGGCTTTTAGTATTGGACAGGCATCCCAAGCAATTCAGCAATTCGAAACGCA 1096
QY      361 ArgGlyAlaAlaIleValGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
Db      1097 AGAGGAGCAGCTTATGAATAATCTTCAAGATAATTTGACAAATAACCAAGCATTTGACAGCTAT 1156
QY      381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db      1157 TCAAGAGTGGACATAAACCAGATAATATTAAAGGAAATTTGGAATTCAAAATGTTTCAAC 1216
QY      401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db      1217 TTCAGTTACCCCTCTCGAAAGAAGACTTAAGATCTTTAAAGGCTCTCAACCTGAAGTTCAG 1276

```


Db 3437 GTCTGATCATCATGAGAGATTGTGCGAGCAGCCCAAGGAGGCCAACATACACCACTTCATC 3496

Qy 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180

Db 3497 GAGACACTCCCTGAGAAATACACACAGAGTAGGAGAGCAAGGAAACCCAGCTCTCTGGT 3556

Qy 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200

Db 3557 GGCGAAGAACAGCGCATTCGCCATAGCTCGCGCTCTGTGTAGACAGCCCTCATATTTTGCTT 3616

Qy 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGlnAlaLeu 1220

Db 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGGTTGTCCAGAGAGCCCTG 3676

Qy 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240

Db 3677 GACAAAGCCAGAGAGCGCGCACTTCATGTGATCGCCACCGCTTGTCCACCATCCAG 3736

Qy 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260

Db 3737 AATGCAGATTTAATAGTGTGTTCAGAAATGCGAAAGTCAAGGAGCATGGCCACATCAA 3796

Qy 1261 GlnLeuLeuAlaGlnLysGlyIleIlePheSerMetValSerValGlnAlaGlyAlaLys 1280

Db 3797 CAGCTGCTGCCCGAGAAAGGCATCTATTTTCCATGGTCAGTGTCCAGGCTGGAGCAAG 3856

Qy 1281 Arg 1281

Db 3857 CGC 3859

RESULT 2

AAD03505

ID AAD03505 standard; cdna; 4279 BP.

XX

AC AAD03505;

XX

DT 13-JUN-2001 (first entry)

XX

DE Dog P-glycoprotein (PGP) allelic variant (Genotype B) cdna.

XX

KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;

KW MDRI; drug bioavailability; transgenic animal; genetic model; ss.

XX

OS Canis familiaris.

XX

FH Key

FT Location/Qualifiers

FT CDS

FT 17..3862

FT /*tag= a

FT /product= "dog P-glycoprotein (PGP) allelic variant

FT (Genotype B) protein"

FT replace (91, T)

FT /*tag= b

FT replace (607, C)

FT /*tag= c

XX

PN W0200123540-A2.

XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-US26767.

XX

PR 28-SEP-1999; 99US-0156510.

XX

PA (GENT-) GENTEST CORP.

XX

PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;

XX

XX WPI; 2001-235373/24.

DR P-PSDB; AA000309.

XX

XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful

PT for determining the bioavailability of drugs and for screening for dog

PT PGP inhibitors -

XX Claim 9; Page 93-99; 111pp; English.

XX

CC The invention relates to dog P-glycoprotein (PGP) also referred

CC as multidrug transporter (MDRI) and nucleic acids encoding them.

CC The invention also includes fragments and biologically functional

CC variants of dog P-glycoprotein, PGP and their nucleic acids are

CC useful for determining the bioavailability of drugs and for

CC screening PGP inhibitors. They are useful for the diagnosis and

CC treatment of conditions characterised by PGP activity, by

CC reducing or increasing PGP activity in a cell. PGP nucleic acids

CC are used as oligonucleotide probes. Complements of PGP nucleic

CC acids are useful as antisense oligonucleotides, to induce a PGP

CC 'knockout' phenotype. They are used to prepare a non-human

CC transgenic animal, which are valuable as genetic models for

CC human diseases.

CC The present sequence is dog P-glycoprotein (PGP) allelic variant

CC (Genotype B) cdna. The PGP enzyme functions as an efflux pump

CC exporting small molecules across the cell membrane. This enzyme

CC is a member of the ABC transporter family.

XX

SQ Sequence 4279 BP; 1296 A; 833 C; 1008 G; 1142 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 4279

Score: 6466.00 Matches: 1279

Percent Similarity: 100.00% Conservative: 2

Best Local Similarity: 99.84% Mismatches: 0

Query Match: 99.89% Indels: 0

DB: 22 Gaps: 0

US-09-672-725C-27 (1-1281) x AAD03505 (1-4279)

Qy 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20

Db 17 ATGGATCCTGAAGGAGCGCGTAAGGGAGTGCAGAGAGAACTTCTGAAAATGGGCAA 76

Qy 21 LysSerLysLysLysGluLysLysGluLysProThrValSerThrPheAlaMetPhe 40

Db 77 AAAAGTAAAAAAGAGAGAGAAAGAAAGAAACCACTGTCCAGACGTTTGCATGTTT 136

Qy 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 60

Db 137 CGCTATTCAATTTGGCTTGATAGTTGTATATGTTGGGGACAATGGCTGCATCATC 196

Qy 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80

Db 197 CATGGAGCTGCACCTCCCTCTCATGATGCTGTTTTTTGGAAACATGACAGATAGCTT 256

Qy 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100

Db 257 AATCAGGAATTTCAAGAAACAAAACCTTTTCCAGTTTATATTAATGAAGTATTACGAC 316

Qy 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120

Db 317 AATACACAACTTTCATCAACCATCTGGAGGAGAAATGACCAGTATGCTATTATTAC 376

Qy 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140

Db 377 AGTGGATCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 436

Qy 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160

Db 437 GCAGCAGGAGACAGACTACTCAAAATTAGAANAACATTTTTCATGCTATCATGCGACAG 496

Qy 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180

Db 497 GAGATTGCTGTTTGAGTGCATGACGTTGGGAGCTTAACACCCGCTGCACAGAGCAT 556

Qy 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200

Db 557 GTCTCCAAAATCAATGAGGAATTTGGCGACAAAATTTGGAATGTTTCTTCAATCAATAGCA 616

QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
Db ACATTTTTCACCGGTTTATAGTGGGTATACACGTGGTGGAGCTAACCCCTGTGATT 676
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
Db TTGGCCCATCAGCCCTGTGTTCTGGACTTTCAGCCGCCATCTGGCGAAGATATCTATCTCA 736
QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
Db TTACTGATAAAGAACTCTGGCCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 796
QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
Db GCAGCAATCAGAACTGTGATTCCTTTGGAGGACAAAGAAAGAACTTGAAGAGTACAAC 856
QY 281 LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
Db AAAAATTTAGAAAGAGCTTAAAGGAATTTGGGATAAGAAAGCTATCAGCGCCAACTTCT 916
QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
Db ATTGGTCGCGCTTCTTAFTGATCATGATCATATGCTCTGGCTTCTGGTATGGGACC 976
QY 321 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPheSerVal 340
Db TCCTTGGTCTCTCCAGTGAATATCTATTTGGACAAGTACTCACTGCTCTTTTCGTA 1036
QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
Db TTAATTTGGGGCTTTTAGTATTTGGACAGCATCCCAAGCATTTGAAGCATTTGCAAAAGCA 1096
QY 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
Db AGAGGACAGCTTATGAAATCTTCAAGATAATTGACAAATAACCAACCATTTGACAGTAT 1156
QY 381 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
Db TCGAAGAGTGACATAAACAGATAAATATTAAAGGAAATTTGGAATTTCAAAAATGTTTAC 1216
QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
Db TTCAATTTACCTTCTCAAAAGAAGTTAAGATCTTAAAGGGCTCTCAACCTCGAAGGTTTCA 1276
QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db AGTGGCAGACAGTGGCGCTGGTTGGACACAGTGGCTGGCGGAAGACACACCGTGCAG 1336
QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
Db CTGATGCAGAGGCTCTATGACCCACAGATGGCATGGCTGTGTATTGATGGACAGGACATT 1396
QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db AGGACCAATAATGTAAGGCATCTTCGGGAATTTACGTGGTGTGAGTCAAGGACCTGTG 1456
QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db TTGTTTCCACCACGATAGCTGAAACATTCGCTATGGCGCGCAAAATGTCACCATGGAT 1516
QY 501 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db GAGATTGAGAAAGCTGTTAAGGAAGCCAATTCCTATGATTTTATCATGAAACTTACCTAAT 1576
QY 521 LysPheAspThrLeuValGlyArgGlyAlaGlnLeuSerGlyGlnLysGlnArg 540
Db AATTTTGACACTCTGGTTGGAGAGAGGGGCCAGCTAGTGGTGGAGAGCAACAGAGA 1636
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
Db ATCGCCATTCTCGGCGCTGTTCGCAACCCCAAGATCTTCTGCTGGATGAGGCAAGC 1696
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580

Db TCAGCTCTGCACACTGAAAGTGAAGCAGTGTTCAGGTGCCCTGGATAAGGCCAGAAA 1756
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db GGCGGACTACCATTTGTGATAGCTCATCGTTTGTCTACAGTTCTGTAATGCCGATGTCATT 1816
QY 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db GCTGGTTTGTGATGAGTGCATTGTGGAGAAAGAAATCATGATGAACCTCATGAAGAG 1876
QY 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db AAGGCATTTTACTTCAAACTTGTCAATGCACACAGAGAAATGAATTGAGTTAGAA 1936
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db AATGCCACTGCTGAATCCAAAAGTGAAGTGCCTTGGAAATGTCTCCAAAAGATTCA 1996
QY 661 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
Db GGTTCAGTTTATAAAGAAAGATCAACTCGCAGGAGTATACATGCACCAAGGCCAA 2056
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
Db GACAAAAGCTTGTACAAAAGAGGACTTGAATGACAATGTACCTCCAGTTCTCTCTGG 2116
QY 701 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db AGGATTTCTGAAGCTGAACCTCACTGAATGGCTTATTTTGTGTTGTTGTTGTTGCT 2176
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db ATTATTAACGGAGGCTCGAACCGCATTTTCAATAATATTTTCAAGGATTTATAGGGATC 2236
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db TTTFACCGAGATGAGGATCCTGAAACAAACGACAGAAATAGTAACATGTTTCTGTATTG 2296
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db TTTCTAGTCTCTGGAAATTTATTTCTTTATATACATTTTCTCCAGGGCTTCACATTTGGC 2356
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db AAAGCTGGGAGATCTCTACTAAGCGCTTCGATACATGTTTTCAGATCCATGCTGAGA 2416
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db CAGGATGTCAGCTGGTTTGTATGACCCCTAAACACACCACTGGAGCATTCACAACAGGCTT 2476
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db GCCAATGATGCGGCTCAAGTTAAAGGGCTATAGTTTCCAGGCTTGTCTCATTAACCCAG 2536
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
Db ATATAGCAAAATCTTGGGACAGCATTTATATCTTAATCTATGTTTGGCAATTAACA 2596
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db CTTTTACTCTTAGCAATTTGATCCCATCATGTCATAGCAGGAGTTGTTGAAATGAAATG 2656
QY 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGlyAlaGlyLysIleAlaThr 900
Db TTGCTCGCAACAGCACTGAAAGATAAGAAAGAGCTAGAAAGAGGCTGGGAAGATTGCTACA 2716
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlyTyr 920
Db GAAGCCATCGAAAACCTTCGAACTGTTGTTCTTTGACTCGGGAGCAGAAAGTTTGAATAC 2776
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940

Db 2777 ATGTATGCACAGAGTTTGAAGTACCATACAGAAACTTTTGAGGAAAGCACACATCTTTC 2836

Qy 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960

Db 2837 GGGGTCTCATTTCTATCACCCAGGAATGATATTTTCCTATGCTGGCTGTTCCGG 2896

Qy 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 980

Db 2897 TTTGGTGCCCTACCTGGTGGCAATGAGTTTCATGAATTTTCAGGATGTTCTTTGGTATTC 2956

Qy 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000

Db 2957 TCAGCTATTGCTTTGGTGCCATGGCAGTGGGCGAGTTCAGTTCAATTTGCTCCTGACTAT 3016

Qy 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleGluLysSerProLeuIle 1020

Db 3017 GCCAAAGCCAAAGTATCAGCAGCCACGTCATCATCATGATCATTTGAAAAAGCCCTCTGATT 3076

Qy 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040

Db 3077 GACAGCTACAGCCCTCAGGCCCTCAGCCCAANTACGTTGGAAGGAATGTGACATTTAAT 3136

Qy 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060

Db 3137 GAGGTCGTGTTCAACTATCCACTCGACACAGACATCCCGCTGCTCCAGGGGCTGAGCCCTC 3196

Qy 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080

Db 3197 GAGGTGAAGAGGCCAGAGCGTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGACACACA 3256

Qy 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100

Db 3257 GTTGCTCAGCTCCTAGAGCGCTTCTATGACCCCTTGCTGCTGCTTAATGATGCG 3316

Qy 1101 LysGluLeuLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120

Db 3317 AAGAGATAAAGCACCTGAATGTCCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAG 3376

Qy 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140

Db 3377 GAGCCCATCCTGTTGACTGAGCATTTGCCGAGAACAATGCCATATGGAGACAACACCCGG 3436

Qy 1141 ValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1160

Db 3437 GTCGTATCACATGAAGAGATTATGCAGGCAGCCAAAGGAGGCCAATACACCACTTCATC 3496

Qy 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180

Db 3497 GAGACACTCCCTGAGAATAACACACAGAGTAGGAGACAAAGGACCCAGCTCTCTGGT 3556

Qy 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200

Db 3557 GCCAGAAACAGCGCATTCGATAGCTCGCGCTCTGTTAGACAGCCTCATATTTTGCTT 3616

Qy 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220

Db 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTGTCCAAGAGCCCTG 3676

Qy 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240

Db 3677 GACAAAGCCAGAGAGGCCGCGACCTGCATTTGTATCGCCACCGCTGTCTCCACCATCCAG 3736

Qy 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260

Db 3737 AATGCAGATTTTAATAGTGGTGTTCAGAAATGGCAAGTCAAGGAGCATGGCACACATCAA 3796

Qy 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280

Db 3797 CAGCTGCTGGCCCAAGAGGCATCTATTTTCCATGTGTGCTCCAGGCTGGAGCAAG 3856

Qy 1281 Arg 1281

Db 3857 CGC 3859

RESULT 3

AA03504

ID AA03504 standard; cDNA; 4279 BP.

XX

AC AA03504;

XX

DT 13-JUN-2001 (first entry)

XX

DE Dog P-glycoprotein (PGP) allelic variant (Genotype A) cDNA.

XX

XX Dog; P-glycoprotein allelic variant; multidrug transporter; MDR1;

KW drug bioavailability; transgenic animal; genetic model; ss.

XX

OS Canis familiaris.

XX

XX Key Location/Qualifiers

FT CDS 17..3862

FT

FT /product= "Dog P-glycoprotein (PGP) allelic variant

FT (Genotype A) protein"

FT replace (607, C)

FT allele

FT /*tag= b

XX

PN WO200123540-A2.

XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-US26767.

XX

PR 28-SEP-1999; 99US-0156510.

XX

PA (GENT-) GENTEST CORP.

XX

PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;

XX

DR WPI; 2001-235373/24.

DR P-PSDB; AAE00308.

XX

XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful

PT for determining the bioavailability of drugs and for screening for dog

PT PGP inhibitors -

XX

PS Claim 9; Page 85-90; 11pp; English.

XX

CC The invention relates to dog P-glycoprotein (PGP) also referred

CC as multidrug transporter (MDR1) and nucleic acids encoding them.

CC The invention also includes fragments and biologically functional

CC variants of dog P-glycoprotein, PGP and their nucleic acids are

CC useful for determining the bioavailability of drugs and for

CC screening PGP inhibitors. They are useful for the diagnosis and

CC treatment of conditions characterised by PGP activity, by

CC reducing or increasing PGP activity in a cell. PGP nucleic acids

CC are used as oligonucleotide probes. Complements of PGP nucleic

CC acids are useful as antisense oligonucleotides, to induce a PGP

CC 'knockout' phenotype. They are used to prepare a non-human

CC transgenic animal, which are valuable as genetic models for

CC human diseases.

CC The present sequence is dog P-glycoprotein (PGP) allelic variant

CC (Genotype A) cDNA. The PGP enzyme functions as an efflux pump

CC exporting small molecules across the cell membrane. This enzyme

CC is a member of the ABC transporter family.

XX

SQ Sequence 4279 BP; 1295 A; 833 C; 1008 G; 1143 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 4279

Score: 6461.00 Matches: 1278

Percent Similarity: 99.92% Conservative: 2

Best Local Similarity: 99.77% Mismatches: 1

Query Match: 99.81% Indels: 0

DB: 22 Gaps: 0

US-09-672-725c-27 (1-1281) x AAD03504 (1-4279)

QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPhetrLysMetGlyLys 20
DB ATGGATCCTGAAGGAGCCCTAAGGGAGTGCAGAGAACTTCTGGAAATAGGCCAAA 76
QY 21 LysSerLysLysGlyLysLysGlyLysLysProThrValSerThrPheAlaMetPhe 40
DB AAAAGTAAAAAATGAGAAGAAAGAAAGAAACCAACTGTCAGCACGTTTGCATGTTT 136
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 60
DB CCTATTTCAATTTGGCTTGATAGTGTATATGTTGGTGGGACAAATGGCTGCCATCATC 196
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
DB CATGGAGCTGCACCTCCCTCTCATGATGCTGTTTTTGGAAACATGACAGATAGCTTTGCA 256
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
DB AATGCAGGAATTTCAAGAAACAAAACCTTTTCCAGTTATAATTAATGAAGATATTAGGAAC 316
QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120
DB AATACACAAACATTTTCATCAACCATCTGGAGGAGAAATGACCAGCATATGCTATTATTAC 376
QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
DB AGTGGGATCGGTGCTGGCGTCTGCTGCTTACATCCAGGTTTCAATCTGGTGCTG 436
QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
DB GCAGCAGGAAGACAGATACTCAAAATTAGAAACAAATTTTTTCATGCTATCATCGACAG 496
QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
DB GAGATTGGCTGGTTGAGTGCATGACGCTTGGGAGGCTTAACACCGGCTCACAGAGCAT 556
QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
DB GTCTCCAAAATCAATGAAGGAATTTGGCGACAAAATTTGGAATGTTCTTCAATCAATAGCA 616
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuValIle 220
DB ACATTTTTTCACCGGTTTTATAGTGGGTTTTACACGTGGTTGGAAGCTAACCCITGTGATT 676
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
DB TTGGCCATCACCCCTGTTCTTGGACTTTTCAGCCGCCATCTGGGCAAGATACTATCTCA 736
QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu 260
DB TTTACTGATAAAGAACTTTGGCCCTATGCAAAAGCTGGACGATAGCTGAAGAAGTCTTA 796
QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
DB GCAGCAATCAGACTGTGATGCTTTGGAGGACAAAAGAAAGAACTTGAAGAGTACAAC 856
QY 281 LysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
DB AAAAATTTAGAAGAAAGCTAAAGGAATTTGGGATAAAGAAAGCTATCAGGGCAACATTTCT 916
QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
DB ATTTGGTCCGCTTTCTATTATGATCATGCAATCATATGCTCTGGCTTCTGGTATGGACC 976
QY 321 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 340
DB TCCTTGGTCTCTCCAGTGAATATCTATTGGCAAGTACTACTGCTCTCTTTCTTTCTGTA 1036
QY 341 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 360
DB TTAATTTGGGCTTTTAGTATTGGACAGGCATCCCAAGCATTTGAAGCATTTTGCAACGCA 1096

QY 361 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 380
DB AGAGGACAGCTTATGAATCTTCAAGATAATTGACAAATAAACCAACGATTCAGACGAT 1156
QY 381 SerLysSerGlyHisLysLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 400
DB TCGAAGAGTGGACATAAACCGATATATTAAGGAAATTTGGAATTCAAAAATGTTTAC 1216
QY 401 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 420
DB TTCAGTTACCCCTCTCGAAAAGAAAGTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTAC 1276
QY 421 SerGlyGlnThrValAlaLeuValGlyAspSerGlyCysGlyLysSerThrThrValGln 440
DB AGTGGGACAGACAGTGGCGCTGGTTGGGAACAGTGGCTGCGGGAAGACACGACCGTCAG 1336
QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 460
DB CTGATGCAGAGGCTCTATGACCCACAGATGGCATGGTCTGTATGATGGACAGGACATT 1396
QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
DB AGGACCATAAATGTAAGGCATCTTCGGGAAATTTACTGGTGTGCTGAGTCAGGACCGCTG 1456
QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
DB TTGTTTCCACCACGATAGTAGAAAACATTCGCTATGGCGCGAAAATGTCACCATGGAT 1516
QY 501 GluIleLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
DB GAGATTGAGAAAGCTGTTAAGGAAGCCATGCCCTATGATTTTATCATGAACATCCTTAAT 1576
QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
DB AAATTTGACACTCTGGTTGGAGAGAGAGGGGCCAGCTGAGTGGTGGCAGAAACAGAGA 1636
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
DB ATCGCCATTGCTCGGGCCCTGGTTTCGAACCCCAAGATTTCTTCTGCTGATGAGGCAAG 1696
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
DB TCAGCTCTGGACACTGAAGATGAAGCAGTGGTTACAGTGGCCCTGGATGAAGCCAGAAA 1756
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
DB GGCCGAGCTACCATTTGTATAGCTCATCGTTTTGTCTACAGTTTCGTAAATGCCGATGCTAT 1816
QY 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
DB GCTGGTTTTGATGATGGAGTCAATTTGGAGAAAGAAATCATGATGAATCATGAAAGAG 1876
QY 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
DB AAGGGCATTTACTTCAAACTTGTCAATGCAGACAGAGGAATGAAATTCAGTTAGAA 1936
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
DB AATGCCACTGGTGAATCCAAAAGTAAAAGTATGCTTGGAAATGCTCTCCAAAAGATTC 1996
QY 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
DB GGGTCAGTTTTTAATAAAAGAAAGATCAACTCCGAGGAGTATACATGACCAACAGGCCAA 2056
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
DB GACAGAAACCTTGGTACAAAAGAGGACTTGAATGAGATGATACCTCCAGTTTCTTCTCG 2116
QY 701 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
DB AGGATTCTGAAGCTGAAGTCAACTCAACTGAATGGCCCTTATTTTGGTGGTATATTTTGTCT 2176

```
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerArgIleIleGlyIle 740
DB 2177 ATTATAAACGGAGCCGTCACACGAGCATTTCAATATATTTTCAAGGATATATAGGATC 2236
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
DB 2237 TTTACCCGAGATGAGGATCTCGAAACAAACGACAGATAGTAACATGTTTCTGTATTG 2296
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
DB 2297 TTTCTAGCTCTGGAATATTTCTTTTATATACATTTTCTCCAGGGCTTCACATTTGGC 2356
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
DB 2357 AAAGCTGGGAGATCCCTACTAAGCGCTTCGATACATAGTTTTCAGATCCATGCTGAGA 2416
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
DB 2417 CAGGATGTCAGCTGTTTGTATGACCCCTAAACACACCACTGGAGCATTTGACACCGAGCTT 2476
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
DB 2477 GCCAATATGCGGTCGAAGTAAAGGGCTATAGGTTCCAGGGCTTGCTGTCAITATCCAG 2536
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThr 860
DB 2537 AATATACCAATCTTTGGACAGGCAATATATATCTTATCTATGTTGGCAATTAACA 2596
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
DB 2597 CTTTACTCTTAGCAATTTGACCATCATTTGCAATAGCAGGAGTTGTTGAATGAAATG 2656
QY 881 LeuSerGlyGlnAlaLeuLysAspLysGluLeuGluAlaGlyLysIleAlaThr 900
DB 2657 TTGCTGGACAAGCACTGAAAGATAAGAAAGAGCTAGAAGAGCTGGGAAGATGCTACA 2716
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
DB 2717 GAAGCCATCGAAACTTCGGAACCTGTTGTTCTTTGACTCGGGAGCAGAAGTTTGAATAC 2776
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
DB 2777 ATGATGCACAGAGTTTGCAGTAGTACCATACAGAAACCTTTTGAGGAAGACACATCTTC 2836
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
DB 2837 GGGGTCTCATTTCTATCACCCAGCAATGATGATTTTTCCTATGCTGGCTGTTTCCGG 2896
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
DB 2897 TTTGGTGCCTACTTGGTGGCAATGAGTTCATGAACCTTTCAGGATGTTCTTTTGGTATTC 2956
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
DB 2957 TCAGCTATTTGCTTTGGTGCCATGGCAGTGGGCGAGGTCAGTTTCATTTGCTCCTGACTAT 3016
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleLeuGluLysSerProLeuIle 1020
DB 3017 GCCAAGCCCAAGTATACGACGCCACCTCATCATGATCATTTGAAAGAGCCCTCTGATT 3076
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
DB 3077 GACAGTACAGCCCTCAGCGGCTCAAGCCAAATAGTTTGAAGGAAATGTGACATTTAAT 3136
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
DB 3137 GAGGTGCGTGTCAACTATCCCACTCGACACAGATCCCGGCTCCAGGGGCTGAGCCCTC 3196
QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
DB 3197 GAGGTGAAGAGGGCCAGACGCTGCGCCCTCGTAGTAGCAGTGGCTGTGGGAAGACACA 3256
QY 1081 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1100
```

```
DB 3257 GTTGTTCAGCTCCTAGAGCGCTTCTATGACCCCTTGGCTGGTTCAGTCTAATTGATGGC 3316
QY 1101 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1120
DB 3317 AAAGAGATAAAGCACCTGAATGTCAGTGGCTCCGAGCACACCTGGGCATCGTCTCTCAG 3376
QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1140
DB 3377 GAGCCCATCTGTTTGACTGCAGCATTTGCCGAGAACATTTGCTATGGAGACAACAGCCGG 3436
QY 1141 ValValSerHisGluGluIleValGlnAlaLysGluAlaAsnIleHisHisPheIle 1160
DB 3437 GTCTGATCACATGAAGAGATTATGCAGCAGCCAAAGGAGGCCAATACACCACTTCAIC 3496
QY 1161 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
DB 3497 GAGACACTCTCCCTCAGAAATATACACACAGAGTAGGAGACAAGGAACCCAGCTCTCTGTT 3556
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
DB 3557 GCCAGAAACAGCGCATTTGCCATAGCTCGCGCTCTTTGTAGACAGCCTCATATTTTGGTT 3616
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
DB 3617 TTGGATGAAGCTACATCATAGCTCTGGATCAGAAAGTGAAGAGTTGTCCAGAAGCCCTG 3676
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
DB 3677 GACAAGCCAGAGAGCGCGCACCTGCTGATGTGATCGCCACCGCTTGTCCACCATCCAG 3736
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
DB 3737 AATGCAGATTTAATAGTGTGTTTTCAGATGTCGAAAGTCAAGGAGCATGGCAGACATCAA 3796
QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
DB 3797 CAGCTGCTGGCCCAAGAGGCACTATTTTTCATGGTTCAGTGTCCAGGCTGGAGCAAG 3856
QY 1281 Arg 1281
DB 3857 CGC 3859
RESULT 4
AAD03488
ID AAD03488 standard; cDNA; 4279 BP.
XX
AC AAD03488;
XX
DT 13-JUN-2001 (first entry)
XX
DE Dog P-glycoprotein (PGP) cDNA #1.
XX
DE Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW drug bioavailability; transgenic animal; genetic model; ss.
XX
OS Canis familiaris.
XX
FH Key
FT CDS 17..3862
FT /*tag= a
FT /product= "Dog P-glycoprotein (PGP) #1"
XX
PN W0200123540-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26767.
XX
XX 28-SEP-1999; 99US-0156510.
XX
PA (GENT-) GENTEST CORP.
```


QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArg 540
|||||
Db 1577 AATTTGACACTCTGGTTGGAGAGAGAGGGCCACCTGAGTGGTGACAGAAACAGAGA 1636
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 560
|||||
Db 1637 ATGCCATTCTCGGGCCCTGGTTCGCAACCCCAAGATCTTCTGCTGGATGAGGCAAG 1696
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
|||||
Db 1697 TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGGTGGCCCTGGATGAAGCCAGAAA 1756
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
|||||
Db 1757 GCCCGACTACCATTTGATAGTCACTCGTTTGTCTACAGTTGCTAATCCGATGTCATT 1816
QY 601 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 620
|||||
Db 1817 GCTGGTTTGTATGATGGAGTCATTGTGGAGAAAGGAAATCATGATGAACCTCATGAAGAG 1876
QY 621 LysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
|||||
Db 1877 AAGGCAATTTACTTCAAACTTTGCACAAATGCACAGAGAGAAATGAATAGATTAGTTAGAA 1936
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
|||||
Db 1937 AATGCCACTGGTGAATCCAAAGTGAAGTGATGCTTGGAAATGTCTCCAAAGATTCA 1996
QY 661 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 680
|||||
Db 1997 GGGTCCAGTTTAAATAAAGAGATCAACTCGCAGGAGTATATCATGCACACAAAGGCCAA 2056
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrp 700
|||||
Db 2057 GACAGAAGCTTGGTACAAAGAGGACTTCAATGAGNATGACTCCAGTTTCTCTCTG 2116
QY 701 ArgIleLeuLysLeuAsnSerThrGluTrpProThrPheValValGlyIlePheCysAla 720
|||||
Db 2117 AGGATTCTGAAGCTGAACCTCAACTGAATGGCTTATTTTGGTTGGTATATTTTGTGCT 2176
QY 721 IleIleAsnGlyLysLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
|||||
Db 2177 ATTATAACGGAGGCGCTTCAACACAGCATTTTCAATAATATTTTCAAGGATATAGGGATC 2236
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
|||||
Db 2237 TTTACCCGAGATGAGGATCTCGAACAACAAACGACAGAAATAGTAACATGTTTCTGTATTG 2296
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
|||||
Db 2297 TTTCTAGTCTCTGGAATATTTCTTTTATTACATTTTCTCCAGGGCTTCACATTTGGC 2356
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgThrMetValPheArgSerMetLeuArg 800
|||||
Db 2357 AAAGCTGGGGAGATCCCTACCTAACCGCTTCGATACATGTTTTCAGATCCATGCTGAGA 2416
QY 801 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
|||||
Db 2417 CAGGATGTCAGCTGGTTTGTATGACCCCTAAAAACACCACTGGAGCATTGACAAACGAGCTT 2476
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
|||||
Db 2477 GCCAATGATGGCGCTCAAGTTTAAAGGGGCTATAGTTTCCAGGCTGTGTCATTACCCAG 2536
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleThrGlyTrpGlnLeuThr 860
|||||
Db 2537 AATATGCAATCTTGGACAGGCAATTTATATATCTTATCTATGTTGGCAATTAACA 2596
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
|||||
Db 2597 CTTTCTACTTCTAGCAATTTGATCCCATCATTTGCAATAGCAGGAGTTGTGAAATGAAATG 2656
QY 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
|||||

Db 2657 TTGCTGCGAAGCACTGAAAGATAAGAAAGAGCTAAGAGAGCTGGAGAGATTGCTACA 2716
QY 901 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTrp 920
|||||
Db 2717 GAAGCATCGAAACTCCGAACCTGTTGTTCTTTGACTCGGAGCAGAACTTTGAATAC 2776
QY 921 MetTrpAlaGlnSerLeuGlnValProTrpArgAsnSerLeuArgLysAlaHisIlePhe 940
|||||
Db 2777 ATGTATGCACAGAGTTTCAAAGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTC 2836
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTrpPheSerTrpAlaGlyCysPheArg 960
|||||
Db 2837 GGGGTCTCATTTCTATCACCAGCAATGATGATATTTTCTATGCTGGCTGTTCCGG 2896
QY 961 PheGlyAlaTrpLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 980
|||||
Db 2897 TTTGGTGGCTACTTGGTGCAAATGAGTTTCATGAACTTTCAGGATGTTCTTTTGGTATTC 2956
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTrp 1000
|||||
Db 2957 TCAGCTATTTGTTCTTTGGTGCCATGGCAGTGGGCGAGGTCACTTCAATTTGCTCCTGACTAT 3016
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
|||||
Db 3017 GCCAAGCCAAAGTATCAGACCCCGCTCATCATCATCATCATTTGAAAAAGCCCTCGATT 3076
QY 1021 AspSerTrpSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
|||||
Db 3077 GACAGCTACAGCCCTCACGGCTCAAGCCAAATACGTTTGAAGGAAATGTGACATTTAAT 3136
QY 1041 GluValValPheAsnTrpProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
|||||
Db 3137 GAGGTCGTGTTCAACTATCCCATCGACACAGACATCCCGTCTCCAGGGGCTGAGCCCTC 3196
QY 1061 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1080
|||||
Db 3197 GAGGTGAAGAGGGCCAGACGCTGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGAGCACA 3256
QY 1081 ValValGlnLeuLeuGluArgPheTrpAspProLeuAlaGlySerValLeuIleAspGly 1100
|||||
Db 3257 GTTGTTCAGCTCCTAGAGCGCTTCTATGACCCCTTGGTGGTTCAGTCTAATTTGATGGC 3316
QY 1101 LysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSerGln 1120
|||||
Db 3317 AAAGAGATAAGACCTGAATGTCAGTGGCTCCGAGCACACCTGGCATCTGCTCTCAG 3376
QY 1121 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTrpGlyAspAsnSerArg 1140
|||||
Db 3377 GAGCCATCTCTTTGACTGACATTCGCGAGAACATTTGCCATATGGAGACAACAGCCGG 3436
QY 1141 ValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1160
|||||
Db 3437 GTCGTATCATCATGAAGAGATTATGACGAGCGCAAGAGGGCCAAACATACACCATTTCAATC 3496
QY 1161 GluThrLeuProGluLysTrpAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1180
|||||
Db 3497 GAGACACTCCCTGGAGAAATACAAACACAGATGAGAGACAAAGAACCCAGCTCTCTGGT 3556
QY 1181 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1200
|||||
Db 3557 GCCCAAAACAGCGCATTTGCCATAGCTCGCGCTCTTGTGTAGACAGCTCATATTTTGTCTT 3616
QY 1201 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1220
|||||
Db 3617 TTGGATTGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAGAACCCCTG 3676
QY 1221 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1240
|||||
Db 3677 GACAAAGCCAGAGAGCGCGACCTCATTTGATCGCCCGCTGTGTCCACCATCCAG 3736
QY 1241 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1260
|||||

Db 3737 AATGAGATTAAATAGTGGTCTTTCAGAAATGGCAAAAGTCAAGGAGCATGGCACACATCAA 3796
QY 1261 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLys 1280
Db 3797 CAGCTGCTGGCCCAAGAAAGCATCTATTTTCCATGCTGCTAGTGTCCAGGCTGGAGCAAG 3856
QY 1281 Arg 1281
Db 3857 CGC 3859
RESULT 5
AAD03489
ID AAD03489 standard; cDNA; 4317 BP.
XX
AC AAD03489;
XX
DT 13-JUN-2001 (first entry)
XX
XX Dog P-glycoprotein (PGP) cDNA #2.
DE Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW drug bioavailability; transgenic animal; genetic model; ss.
XX
XX Canis familiaris.
XX
FH Key Location/Qualifiers
FT CDS 70..3912
FT FT /*tag= a
FT FT /product= "Dog P-glycoprotein (PGP) #2"
XX
PN WO200123540-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26767.
XX
XX 28-SEP-1999; 99US-0156510.
XX
XX (GENT-) GENTEST CORP.
XX
PI Stocker PJ, Steinel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX
XX WPI; 2001-235373/24.
DR P-PSDB; AAE00304.
XX
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
XX Claim 1; Page 66-72; 111pp; English.
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDR1) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) cDNA. The
CC PGP enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.
XX
SQ Sequence 4317 BP; 1293 A; 844 C; 1019 G; 1161 T; 0 other;
Alignment Scores: 0 Length: 4317
Pred. No.:

Score: 6425.50 Matches: 1273
Percent Similarity: 99.69% Conservative: 4
Best Local Similarity: 99.38% Mismatches: 3
Query Match: 99.27% Indels: 1
DB: 22 Gaps: 1
US-09-672-725C-27 (1-1281) x AAD03489 (1-4317)
QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 70 ATGGATCCTGAAGGAGCGCTAAGGGAGTGCAGAGAAGAACTTCTGGAAAATGGCAAA 129
QY 21 LysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
Db 130 AAAAGT---AAAAAGAGAGAAAGAAAGAAACCAACCTGTGAGCAGCTTTCGAATGTT 186
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 60
Db 187 CGCTATTCAAAATTTGGCTTCATAGTTGTATATGTTGGTGGGACAATGGCTGCCATCATC 246
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
Db 247 CATGGAGCTGCACCTCCCTCTCATGATGCTGCTTTTGGAAACATGACAGATAGCTTTCGA 306
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 307 AATGCAGGAATTTCAAGAAACAAACCTTTCCAGTTATATAATTAATCAAGATATTACGAAC 366
QY 101 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 120
Db 367 AATACACACATTTTCATCAACCATCTGGAGGAGGAATGACCACCATGATGCCATATTATAC 426
QY 121 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 140
Db 427 AGTGGGATCGGTGCTGGCTGCTGCTTACATCCAGCTTTCATCTCTGGTGCCTG 486
QY 141 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 160
Db 487 GCAGCAGGAGAGACAGATACTCAAAATTAGAAACAAATTTTTCATGCTATCATCGCAGAG 546
QY 161 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 180
Db 547 GAGATTGGCTGCTTTCACGTGCATCAGCTGGGAGCTTAACACCCGGCTCACAGACGAT 606
QY 181 ValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIleAla 200
Db 607 GTCTCCAAAATCAATGAAGGAATTCGCACAAAGTTGGAATGTTCTTTCAATCAATAGCA 666
QY 201 ThrPhePheThrGlyPheIleValGlyPheThrArgGlyTyrTrpLysLeuThrLeuValIle 220
Db 667 ACATTTTTCACCGGTTTATAGTGGGGTTTACACCTGGTTGGAAGCTAACCTTGTGATT 726
QY 221 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 240
Db 727 TTGGCCATCAGCCCTGCTTCTGGACTTTCAGCCGCCATCTGGGCAAGATACATCTCTCA 786
QY 241 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 260
Db 787 TTTACTGATAAAGAACTCTTGGCCCTATGCAAAAGCTGAGCAGTAGCTGAAGAAAGTCTTA 846
QY 261 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 280
Db 847 GCAGCAATCAGAACTGTGATTCCTTTGGAGGACAAAGAAAGAACTTTGAAAGGTACAAC 906
QY 281 LysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSer 300
Db 907 AAAAATTTAGAAGAGCTAAAGAAATTTGGGATAAAGAAAGCTATCACGGCCCAACATTCT 966
QY 301 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 320
Db 967 ATTGGTCCGCTTTCTTATTGATCATCATATGCTCTGGCTTCTTGTATGGGACC 1026
QY 321 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 340

Db 1027 TCCCTGGCTCTCCAGTGAATATCTACTATTGGACAGGTACTCAGTCTCTCTTTCTGTA 1086
QY 341 LeuileGlyAlaPheSerileGlyGlnAlaSerProSerileGluAlaPheAlaAsnAla 360
Db 1087 TTAATTTGGGGCTTTTAGTATTGGACAGGCATCCCCAAGCATTTGAAGCATTTGCAAAACGCA 1146
QY 361 ArgGlyAlaAlaTyrGluilePheLysilelleAspAsnLysProSerileAspSerTyr 380
Db 1147 AGAGGAGCAGCTTATGAATCTTCAAGATAATTGACAATAATTAACCAACCAAGCATTCGACGCTAT 1206
QY 381 SerLysSerGlyHisLysProAspAsnilleLysGlyAsnLeuGluPheLysAsnValHis 400
Db 1207 TCGAAGCTGGACATTAACACGAGATAATTAAGGGAAATTTGGAAATTCAAAAATGTTCCAC 1266
QY 401 PheSerTyrProSerArgLysGluValLysilleLeuLysGlyLeuAsnLeuLysValGln 420
Db 1267 TTCAGTTACCTTCTCGAAAAGAAAGTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTTCAG 1326
QY 421 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 440
Db 1327 AGTGGCGACAGCTGGCGTGGTGGGAACAGTGGCTGCGGGAAGACGACCGCTGCGAG 1386
QY 441 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysilleAspGlyGlnAspille 460
Db 1387 CTGATCGACAGGCTCTATGACCCACAGATGGCATGGTCTGTATTGATGGACAGGACATT 1446
QY 461 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 480
Db 1447 AGGACCATAAATGTAAGGCATCTTCGGGAAATTTACTTGGTGGTAGTCAGGAGCGCTGFG 1506
QY 481 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 500
Db 1507 TTGTTTCCACCACGATAGCTGAAACATTCGCTATGCGCGGAAATGTCACCATGGAT 1566
QY 501 GluileGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 520
Db 1567 GAGATTGAGAAAGCTGTTAAGGAAGCAATGCTATGATTTATCATGAACCTACCTAAT 1626
QY 521 LysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlnLysGlnArg 540
Db 1627 AAATTTGACACTCTGGTTGGAGAGAGGGGGCCCGTGTAGTGGTGGACAGAAACAGAGA 1686
QY 541 IleAlaIleAlaArgAlaLeuValArgAsnProLysilleLeuLeuLeuAspGluAlaThr 560
Db 1687 ATGCCATTGCTCGGGCCCTGGTTCCGCAACCCCAAGATTCTTCTGCTGGATGAGGCAACG 1746
QY 561 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 580
Db 1747 TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAAGTGGCCCTGGATAAGGCCAGAAA 1806
QY 581 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 600
Db 1807 GCGCGGACTACCATGTGTATAGCTATCGTGTGTCTACAGTTCTGTAATGCCGATGTCAAT 1866
QY 601 AlaGlyPheAspGlyValIleValIleGluLysGlyAsnHisAspGluLeuMetLysGlu 620
Db 1867 GCTGGTTTGTATGATGAGTCATTGTGGAGAAAGGAAATCATGATCAACTCATGAAAGAG 1926
QY 621 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 640
Db 1927 AAGGGCATTTACTTCAAACTTGTCCAAATGCACAGACAGAGGAAATGAAATGGATGTAGAA 1986
QY 641 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 660
Db 1987 ANTCCACTGTGTGAATCCAAAGTGAAGTGAATGATGCCCTTGGAAATGTCTCCAAAGATTCA 2046
QY 661 GlySerSerLeuIleLysArgArgSerThrArgSerIleHisAlaProGlnGlyGln 680
Db 2047 GGGTCCAGTTTAAATAAAAGAAAGATCAACTCGCAGGAGTATACATGCACCACCAAGGCCAA 2106
QY 681 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTyr 700

Db 2107 GACAAAAGCTTGGTACAAAAGAGGACTTGAATGAGAAATGTTCTCTCAGTGTCTCTTCTGG 2166
QY 701 ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 720
Db 2167 AGGATTCGAAAGCTGAAGCTCAACTGAATGGCTTATTTTGGTGGTATATATTTGTGCT 2226
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2227 ATTTAAACGAGGCGCTGCAACACGACATTTTCAATAATATTTTCAAGGATATATAGGATC 2286
QY 741 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 760
Db 2287 TTTACCCGAGATAGGATCTCTGAACAAACACGACAGAAATAGTAACATGTTTCTGTATG 2346
QY 761 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 780
Db 2347 TTTCTAGTCCCTGGAAATATTTCTTTTATTACATTTTCTCCAGGCTTCACATTTGGC 2406
QY 781 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 800
Db 2407 AAAGCTGGGAGATCTCTCACTAAGCGCTTCGATACATGTTTTCAGATCCATGCTCAGA 2466
QY 801 GlnAspValSerTyrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 820
Db 2467 CAGGATGTCAGCTGGTGGTGGACCCCTTAAACACCACTGGAGCATTCGACCAACGAGCTT 2526
QY 821 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 840
Db 2527 GCCAATGATCGGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTGTGTCTATTACCCAG 2586
QY 841 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTyrGlnLeuThr 860
Db 2587 AATATAGCAAAATCTTGGACAGGCATTTATATATCTTAAATCTATGTTGGCAATTAAAC 2646
QY 861 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 880
Db 2647 CTTTACTCTCTAGCAATTTGATCCCATCATTCGAATGAGGAGTGTGTAATGAAATG 2706
QY 881 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 900
Db 2707 TTGCTGGCAAGCACTGAAAGATAAGAAAGAGCTAGAGGAGCTGGGAAAGATTGCTACA 2766
QY 901 GluAlaIleGluAspPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 920
Db 2767 GAAGCCATCGAAACATCCGAACCTGTTCTTGTACTCGGAGCAGCAAGTTGAATAC 2826
QY 921 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 940
Db 2827 ATGTATGCACAGAGTTTCAAGTACCATACAGAAACTCTTTGAGGAAAGCACATCTTC 2886
QY 941 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 960
Db 2887 GGGGTCTCATTTTCTATCACCAGCAATGATGATTTTCTCTATGCTGGCTGTTTCGG 2946
QY 961 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 980
Db 2947 TTTGGTGCCCTACTTGGTGCAAAATGAGTTCATGAACATTCAGGATGTTCTTTGTAATC 3006
QY 981 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 1000
Db 3007 TCAGCTATTGCTTCTTGGTGCCATGGCAGTGGGCGAGGTCAGTTCTATTTCCTCCTGACTAT 3066
QY 1001 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1020
Db 3067 GCCAAGGCCAAAGTATCAGACGCCACCTCATCATGATCATTTGAAAAAGCCCTCTGATT 3126
QY 1021 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1040
Db 3127 GACAGCTACAGCCCTCAGGCCCTCAAGCCAAATACGTTGGAAGGAAATGTGACATTTAAT 3186
QY 1041 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1060
Db 3187 GAGGTGCTGTTCAACTATCTCCACTCGACCAGACATCCCGTCTCCAGGGGCTGAGCCTC 3246

QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
|||||
Db 178 ATCCATGGGCGTGGACTCTCTCATGATGCTGGTGGTGGAGAAATCACAGATATCTTT 237
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
|||||
Db 238 GCAAAATGCAGGA--AATTTAGAAGATCTGATGTCAAAACATCACTAAATAGAAGTGAATC 294
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyraIaTyrr 119
|||||
Db 295 AATGATACAGGGTCTTCATGAAT---CTGGAGGAGACATGACAGATATGCGCTATTAT 351
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaIaTyrrIleGlnValSerPheTrpCys 139
|||||
Db 352 TACAGTGGAAATGGTGCTGGGTGGTGGTGTCTTACATTCAGGTTCATTTTGGTGC 411
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
|||||
Db 412 CTGGCAGCTGGAAGACAAATACACAAATTAGAAAACAGTTTTCATGCTATAATGCGA 471
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
|||||
Db 472 CAGGAGATAGCTGGTTTATGTGCACGATTTGGGGAGCTTAACACCCGACCTTACAGAT 531
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
|||||
Db 532 GATGCTCTTAAGATTAAATGAAGTATTGCTGACAAAATTTGGAATGTCTTTCAGTCAATG 591
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
|||||
Db 592 GCAACATTTTCTACTGGGTATTAGTAGGATTTTACACGTGGTGGAGCTTAACCCCTTGG 651
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleIleIleIleIleLeuSer 239
|||||
Db 652 ATTTTGGCCATCAGTCTCTGTCTGTGACGTGCTGCTGCTGGCCAAAGTACTATCT 711
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrrAlaLysAlaGlyAlaValAlaGluVal 259
|||||
Db 712 TCATTTACTATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTCT 771
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysGlyLeuGluArgTyr 279
|||||
Db 772 TTGGCAGCAATTAAGAACTGTGATGTGATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 831
QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
|||||
Db 832 AACAAAAATTTAGAAAGCTAAAGAAATTTGGGATAAGAAAGCTATTACAGCAATATT 891
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrrAlaSerTyrrAlaLeuAlaPheTrpTyrrGly 319
|||||
Db 892 TCTATAGTGCTGCTTTCTCTGCTGATCTATGCACTTATGCTCTGGCCCTCTGGTATGGG 951
QY 320 ThrSerLeuValLeuSerSerGluTyrrThrIleGlyGlnValLeuThrValPhePheSer 339
|||||
Db 952 ACCACCTTGTCCTCTCAGGGGAATATCTATTGGACAGTACTACTGATTTCTTTCT 1011
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
|||||
Db 1012 GTATTAATTTGGGCTTTTAGTGTGGACAGGCATCTCCAAGCATTTGAAGCATTTGCAAAAT 1071
QY 360 AlaArgGlyAlaAlaTyrrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
|||||
Db 1072 GCAAGAGGACGACGCTTATGAAATCTTCAAGATAATTAAGCAAGTATGACAGC 1131
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
|||||
Db 1132 TATTTCGAAGAGTGGCACAACACAGATAATTAAGGGAATTTGGAATTCAGAAATGTT 1191
QY 400 HisPheSerTyrrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
|||||
Db 1192 CACTTCACTGTTACCCATCTCGAAAAAGAAAGTTAAGATCTTTGAAGGCGCTGAAACCTGAAAGGTG 1251
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439

Db 1252 CAGAGTGGCAGACGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAGAGACACAAGATC 1311
QY 440 GlnLeuMetGlnArgLeuTyrrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
|||||
Db 1312 CAGCTGATGACAGAGGCTCTATGACCCACAGAGGGATGCTCAGTGTGTGATGACAGGAT 1371
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
|||||
Db 1372 ATTAGGACCAATAATGTAAGTTTCTACGGGAAATCATTTGCTGTGTGAGTCAGGAACCT 1431
QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrrGlyArgGluAsnValThrMet 499
|||||
Db 1432 GTATTGTTGGCCACCACGATAGCTGAAACATTCGCTATGCGCGTGAATGTCACCATG 1491
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrrAspPheIleMetLysLeuPro 519
|||||
Db 1492 GATGAGATGGAAGAAGCTGTCAAGAAAGCCATGCTTATGACCTTATCATGAAACCTGCT 1551
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
|||||
Db 1552 CATAAATTTGACACCCCTGGTTGGAGAGAGAGGGGCCAGGTGAGTGGTGGGCGAAGACAG 1611
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
|||||
Db 1612 AGGATCGGCATTTGACGCTGCGCTGCTGCAACCCCAAGATCCTCCTGCTGGATGAGGCC 1671
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
|||||
Db 1672 ACCTCAGCTTGGACACAGAAAGCGAAGCAGTGTTCAGGTGGCTCTGGTAAGCCAGA 1731
QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
|||||
Db 1732 AAAGTCGACGACCACTTGTGATAGCTCATGCTTGTCTACAGTTCGTAATGCTGACGCT 1791
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
|||||
Db 1792 ATCGCTGTTTTCGATGATGAGTCACTTGTGAGAAAGAAATCATGTAACCTCATGAAA 1851
QY 620 GluLysGlyIleTyrrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
|||||
Db 1852 GAGAAAGCAATTTACTTCAAACTTGTCAAAATGCAGACAGCAGGAAATGAAGTTGAATTA 1911
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAlaLeuGluMetSerProLysAsp 659
|||||
Db 1912 GAAATCCAGCTGATGAATCCAAAGTGAATTCATGCTTGGAAATGCTCTCAATGAT 1971
QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
|||||
Db 1972 TCAAGATCCAGTCTAATAAGAAAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAGCC 2031
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
|||||
Db 2032 CAAGACAGAAAGCTTAGTACCAAGAGAGCTCTGGATGAAGATATACCTCCAGTTTCCCTT 2091
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrrPheValValGlyIlePheCys 719
|||||
Db 2092 TGGAGATTAATGAACCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2151
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
|||||
Db 2152 GCCATTAAATGGAGGCTGCAACAGCAGATTTGCAATAATTTTCAAGATATATAGGG 2211
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
|||||
Db 2212 GTTTTACAGAAATTTGATGATCCTGAAACAAACACAGAAATAGTAATCTGTTTCACTA 2271
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
|||||
Db 2272 TTGTTTTCAGCCCTTGGAAATTTATTTTATTTATTTTATTTTATTTTATTTTATTTTATTT 2331
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrrMetValPheArgSerMetLeu 799
|||||

[illegible]

Identifying a stem cell, for treating e.g., muscular dystrophy, myocardial infarction, Parkinson's disease, or neurodegenerative disorders, comprises detecting the expression of an ATP transport protein (BCRP) by a cell -

Disclosure; Page 53-55; 87pp; English.

The invention provides a method of identifying and/or isolating a stem

CC cell that involves detecting the expression of an ATP transport protein
CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
CC comprising stem cells. The isolated stem cells may be used in the
CC treatment of diseases such as muscular dystrophy, degenerative liver
CC disorder, myocardial infarction, Parkinson's disease, degenerative
CC disorders of the brain, and for tissue regeneration or replacement.
CC Haematopoietic cells can be used in bone marrow transplants (e.g., for
CC treatment of leukemia) and for ex vivo gene therapy for treating blood
CC diseases such as sickle cell anemia and thalassemia. The stem cells can
CC also be used as cell targets in gene therapy protocols. The present
CC sequence represents a sequence related to the BCRP for which no relevant
CC information has been provided in the specification.

XX Sequence 3860 BP; 1135 A; 746 C; 957 G; 1022 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 3860
Score: 5866.00 Matches: 1164
Percent Similarity: 95.09% Conservative: 55
Best Local Similarity: 90.80% Mismatches: 59
Query Match: 90.62% Indels: 4
DB: 24 Gaps: 4

US-09-672-725C-27 (1-1281) x ABA94365 (1-3860)

Qy	1	MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly	19
Db	1	ATGATCTTGAAGGGGCCCAATGAGGNGCAAGAAAGAACTTTTAAACTGAAAC	60
Qy	20	LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet	39
Db	61	AATAAAAGT---GAAAAAGATAAGAGAAAGAAACCAACTGTCAGTGTATTTTCAATG	117
Qy	40	PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle	59
Db	118	TTTCGCTATTCAAAATGGCTTGAAGTGTATATGGTGGGAACCTTTGGCGGCATC	177
Qy	60	IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe	79
Db	178	ATCCATGGGGCTGGACTCTCTCATGATGCTGTGTGGAGAAATGACAGATATCTTT	237
Qy	80	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr	99
Db	238	GCAAAATGCAGGA---AATTAGAGATCTGATGCTCAAAACATCACTAATAAGAGTATATC	294
Qy	100	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr	119
Db	295	AATGATACAGGTTCTTCATGAAT---CTGGAGGAGACATGACCAGATATGCTATTAT	351
Qy	120	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys	139
Db	352	TACAGTGGAAATGGTGTGGGGTCTGCTGCTTACATTACAGTTTCATTTTGGTGC	411
Qy	140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159
Db	412	CTGGCAGCTGGAGACAAATACACAAATTAGAAAAACAGTTTTTCATGCTATAATGCCGA	471
Qy	160	GlnGluIleGlyTyrPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179
Db	472	CAGAGATAGGCTGGTTGATGTCACAGATGTTGGGAGGCTTAACACCCGACTACAGAT	531
Qy	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle	199
Db	532	GATGCTCTAAGATTAATGAAGGTATTGGTGACAAAAATGGAATGTTCTTTCAGTCAATG	591
Qy	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219
Db	592	GCAACATTTTTCACCTGGGTTTATAGTAGGATTTACACGTGGTGGAAAGCTAACCCCTGTG	651
Qy	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239
Db	652	ATTTTGGCCATCAGTCTCTTCTTGGACTGTCAGCTGCTGCTGGGCAAGATACTATCT	711

Qy	240	SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal	259
Db	712	TCATTTACTATGAAGAACTCTTAGCGTATGCAGAAAGCTGGACGCTAGCTGAAGAGCT	771
Qy	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluAlaTyr	279
Db	772	TTGGCAGCAATTAGAACTGTGATTCATTTGGAGGACAAAGAAAGAACTTGAAGGTAC	831
Qy	280	AsnLysAsnLeuGluAlaAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299
Db	832	AACAAAAATTAGAAAGAACTAAAGAAATGGGATAAAGAAAGCTATTACAGCAATATT	891
Qy	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319
Db	892	TCATAGGTGCTGCTTCTCTGATCATCTTATGCTCTGGCCTTCTGGTATGGG	951
Qy	320	ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer	339
Db	952	ACCACCTTGGTCTCTCAGGGGAATATTCATTGACAAAGTACTACTGTATCTTTTCT	1011
Qy	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359
Db	1012	GTATTAATGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAAAT	1071
Qy	360	AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379
Db	1072	GCAAGAGAGAGAGCTTATGAATCTTCAAGATAATTGATAATAAGCCCAAGTATTTCACAGC	1131
Qy	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399
Db	1132	TATTCGAAGAGTGGCACAACACAGATAATATTAAAGGAAATTTGGAATTCAGAAATGTT	1191
Qy	400	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419
Db	1192	CACTTCAGTTACCCATCTCGAAAAAGAGTTAAGATCTTGAAGGGCCTGAACCTGAAGGTG	1251
Qy	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal	439
Db	1252	CAGATGGGCGAGACGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGAGCACACAGTC	1311
Qy	440	GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459
Db	1312	CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGGATGGTCAGTGTGTATGGACAGGAT	1371
Qy	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro	479
Db	1372	ATTAGGACCATAAATGTAAGGTTTCTACGGGAAATCATTTGGTGTGGTGTGAGTCAGGAACCT	1431
Qy	480	ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499
Db	1432	GTATTGTTGCCACCCAGCATAGCTGAAACATTCGCTATGGCCGTGAAATGTCAACCATG	1491
Qy	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro	519
Db	1492	GATGAGATTGAGAAAGCTGCTCAAGGAAGCCATATGCTATGATTTATCATGAACTGGCT	1551
Qy	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539
Db	1552	CATAAATTTGACACCCCTGGTTGGAGAGAGGGGGCCAGTTGAGTGTGGCGAGAGACAG	1611
Qy	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559
Db	1612	AGGATGCCCATTTGCAGTGCCTTGGTTCGCAACCCCAAGATCCTCTCTGCTGGATGAGGCC	1671
Qy	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579
Db	1672	ACGTACGCTTGGACACAGAAAGCGAAGCAGTGGTTCAGGTGGCTCTGGATGAAGCCAGA	1731
Qy	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599
Db	1732	AAAGGTGGGACCAACCATTTGTGATAGCTCATCGTTTGTCTACAGTCTGTAATGCTGACGCTC	1791
Qy	600	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619

```
Db 1792 ATCGCTGGTTTCATGATGGAGTCAATTGTGGAGAAAGGAAATCATGATGAACATCATGAAA 1851
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 1852 GAGAAAGGCAATTACTTCAAACTGTGCACATGCGACAGCAGGAGAAATGAAGTTAAATTA 1911
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 1912 GAAATGCGAGCTGATGAATCCAAAAGTGAATTTGATGCTTGGAAATGTCCTCAATGAT 1971
QY 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 1972 TCAAGATCCAGTCTAATAAGAAAAGATCAACTCGTAGGAGTGTCCGTGGATCAACAAGCC 2031
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2032 CAAGACAGAAAGCTTAGTACCAAGAGGCTCGGATGAAGATGAAATACCTCCAGTTTCCTTT 2091
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2092 TGGAGGATTAATGAAGCTAAATTAACCTGAATGGCCCTATTATTGTTGTTGGTGTATTGTT 2151
QY 720 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIlePheSerArgIleIleGly 739
Db 2152 GCCATTAAATGAGAGGCTGCAACAGCATTTGCAATTAATATTTTCAAGATTAATAGGG 2211
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAnSerAsnMetPheSerVal 759
Db 2212 GTTTTACAGAATTGATGATCCTGAAACAAACAGACAGATAGTAACTTTGTTTCTACTA 2271
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2272 TTGTTTTCAGCCCTTGGAAATTAATTTCTTTTATTACATTTTTCCTTCAAGGTTTCACATT 2331
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2332 GGCAAAGCTGGAGAGATCCTCACCAGCGGCTCGGATACATGGTGTTCGATCCATGCTC 2391
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2392 AGACAGGATGTGAGTTGTTGATGACCTAAAACACCACCTGGAGCATTGACTACCAGG 2451
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2452 CTCGCCAATGATGCTGTCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCTGTAATACC 2511
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2512 CAGATATAGCAATCTTGGGACAGGAATAATATATATCTTCATCTATGGTTGGCAACTA 2571
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2572 ACACGTGTACTCTTAGCAATGTACCATCATCTTCATATAGCAGGAGTTGTTGAATGAAA 2631
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluAlaGlyLysIleAla 899
Db 2632 ATGTTGTCTGCACAAGCACTGAAAGATAAGAAAGAACTAGAAAGGTGCTGGGAAGATGCT 2691
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2692 ACTGAAGCAATAGAAAACTTCCGAACCGTTGTTTCTTGTGACTCAGGAGCAGAAAGTTGAA 2751
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2752 CATATGTATGCTCAGATGTTGCAGGTACCATACAGAAACTCTTTGAGAAAGCACACATC 2811
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2812 TTTGGAATTACATTTTCTCCACCAGGCAATGATGATTTTCTTATGCTGGATGTTTC 2871
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
```

```
Db 2872 CGGTTTGGAGCCTACTTTGGTGGCACAATAAACTCATGACCTTTGAGGATGTTCTCTTAGTA 2931
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 2932 TTTTCAGCTGTTGCTTTTGGTGCCATGGCCGTGGGCAAGTCAGTTTCATTTGCTCTCAGC 2991
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 2992 TATGCCAAGCCAAATATATCAGACCCACATCATCATGATCATTTGAAATAACCCCTTTG 3051
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3052 ATTCACAGCTACAGCAGCGAAGGCTTAATGCCGAACACATTGGAAGGAATGTCACATTT 3111
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3112 GGTGAAGTTGTTCAACTATCCACCAGCCGACATCCAGTGTCTCAGGACTGAGC 3171
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3172 CTGGAGGTGAAGAGGCCAGACCTGGCTCTGTGTGGCAGCAGTGGCTGTGGGAAGAGC 3231
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3232 ACAGTGTCTCCAGCTCTCTGGAGCGGTTCTACGACCCCTTGGCAGGGAAGTGTCTGTGAT 3291
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3292 GGCAAGAAATAAAGCAGCTGAATGTTCACTGGCTCGAGCACACCTGGGCATCGTGTC 3351
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3352 CAGGAGCCCATCTCTGTTGACTGCAGCATTCGTGAGAACATTCGCTATGGAGACAACAGC 3411
QY 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3412 CGGTTGTGTACAGGAAGAGATCGTGAGGGCAGCAAGAGGAGCCAAACATACATGCTTC 3471
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3472 ATCGAGTCACTGGCCTAATAAATATAGCACTAAAGTAGGACAAAGGAACCTCAGCTCTCT 3531
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3532 GGTGGCCAGAAACAACGCAATTGCCATAGCTCGTGCCCTGTGTAGACAGCCTCATATTTG 3591
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3592 CTTTGTGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAAGGTTGTCCAAGAAGCC 3651
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3652 CTGGCAAAAGCCAGAGAGGCCGACCTGCATTTGATTGCTCACCCGCTGTCACCATC 3711
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3712 CAGAAATCAGACTTATAGTGGTGTTCAGATGGCAGAGTCAGGAGCATGGCAGCAT 3771
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3772 CAGCAGCTGTGGCACAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 3831
QY 1280 LysArg 1281
Db 3832 AAGCGC 3837
RESULT 8
AAF86127
ID AAF86127 standard; cDNA; 4186 BP.
AC AAF86127;
XX
XX 25-JUN-2001 (first entry)
```

XX Cynomologous monkey P-glycoprotein cDNA.
DE Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL;
XX efflux pump; ss.
KW Macaca fascicularis.
XX
OS
XX
FH Key Location/Qualifiers
XX CDS 100..3942
FT /*tag= a
FT /product= "pgp"
FT /note= "P-glycoprotein"
XX
XX WO2001233565-A1.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26592.
XX
XX 28-SEP-1999; 99US-0156921.
PR 12-OCT-1999; 99US-0158818.
XX
XX (GENT-) GENTEST CORP.
XX
XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
PI
XX WPI: 2001-316136/33.
DR P-PSDB; AAB81064.
XX
XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell -
XX
XX Example 1; Page 51-57; 84pp; English.
XX
XX This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDRL is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents cDNA encoding
CC cynomologous monkey P-glycoprotein.
XX
SQ Sequence 4186 BP; 1226 A; 799 C; 1039 G; 1122 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 4186
Score: 5862.00 Matches: 1163
Percent Similarity: 95.32% Conservative: 59
Best Local Similarity: 90.72% Mismatches: 56
Query Match: 90.56% Indels: 4
DB: 22 Gaps: 4

US-09-672-725C-27 (1-1281) x AAP86127 (1-4186)

QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGlu---LysAsnPhetTrpLysMetGly 19
Db 100 ATGGATCTTGAGGGACCGCAATGGAGGACGACAGAGAAGAACTTTTAAACTGAAC 159
QY 20 LysLysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 160 AATAAAGT---AAAAAGATAAGAGGAAGAAACCAACTGTCAGTGTATTTCATG 216
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 217 TTTCCGTATTCAANTGGCTTCACAAAGTTGTATATGGTGGTGGGAACCTTGGCTGCCATC 276

QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 277 ATCCATGGAGCTGGACTTCTCTCATGATGCTGGTGTGGAGACATGACGGATACCTTT 336
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 337 GCAAAATGCAGGA---AATTTAGGAGATTTAGGAGCTCTGTTGACATAATAGTAGTATC 393
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 394 ACTGATACAGTCCCTCCCTCATGAAT---CTGGAGGAGATATGACCAAGTATGCTTATAT 450
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaIleTyrIleGlnValSerPheTrpCys 139
Db 451 TACAGTGAATTTGGTCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 510
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 511 CTGGCAGCTGGAAGACAAATACACAAATAGAAACAGTTTTTTCATGCTATAATAGCGA 570
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 571 CAGGATAGGCTGGTTCATGTCACGATGTTGGGAGCTTAACACCGCGCTTACAGAT 630
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 631 GATGCTCCAAAGATTAAATGAAGGAATGTTGACAAATTTGGAATGTTCTTTCAGTCAATG 690
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 691 GCAACATTTTCTACTGGGTTTATAGTAGGATTTACACGTTGGAGCTTGAAGCTAACCTTGTG 750
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 751 ATTTGGCCATCAGTCTCTTCTGACGTCAGCTGACGTCAGCTGGGCAAAAGATCTGTCT 810
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 811 TCATTTTACTGATAAAGAACTTTAGCTTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTG 870
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 871 TTGGCAGCAATTAAGAACTGTGATGTCATTTGGAGGACAAAGAAAGAACTCGAAAGGTAC 930
QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 931 AACAAAAATTTAGAAGAAAGCTAAAGAAATTTGGATAAAGAAAGCTATTACAGCCAATAT 990
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 991 TCTATAGGTGCTGCTTCTCTGCTTATCTATGCAATCTATGCTCTGCGCTTCTGGTATGGG 1050
QY 320 ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer 339
Db 1051 ACCACCTTGCTCTCTCAAGGAATATCTATTTGGACAAAGTACTCACTACTATTCTTTCT 1110
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1111 GTATTAATTTGGGCTTTTAGTGTGGACAGGCATCTCCAGCAGCATTTGAAGCATTTGCAAT 1170
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1171 GCAAGAGGACAGCTTTTGAATCTTCAAGTAAATTTGATATAAGCAAGTATTGACAGC 1230
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1231 TATTGGAAGAGTGGGCACAAACCAAGATAATATTAAAGGAAATTTGGAATTCAGAAATGTT 1290
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1291 CACTTCAGTTACCCATCTCGAAAAAGAGTTAAGATCTTTGAAGGGCGCTTGAACCTTGAAGGTG 1350

QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
|||||
Db 1351 CAGAGTGGCAGACGGTGGCCCTGGTTGGAAACACGGCTGGGAAGACGACAAACGGTC 1410
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
|||||
Db 1411 CAGCTGATGCAGAGGCTTATGACCCACAGAGGCGATGGTCAGTGTGTGACAGGAT 1470
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
|||||
Db 1471 ATTAGACCATTAACAGTAAAGTCTTACGGGAAATCATCGGTGGTGGTCAAGAACCT 1530
QY 480 ValLeuPheAlaThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
|||||
Db 1531 GTATTGTTGGCCACCACCATAGCTGAACCATTCGCTATGTCGTGAAGATGTCACCAATG 1590
QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
|||||
Db 1591 GATGAGATTGAGAAAGCTGTCAAGGAAGCCAAATGCTTATCATGAAACTGCCT 1650
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGln 539
|||||
Db 1651 CAGAAATTTGACACCTGGTGGAGAGAGGGGCCCGAGCTGAGTGGTGGCGCAGAGCAG 1710
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
|||||
Db 1711 AGGATCGCCATTGCGACGTGCCTGGTTGCGAACCCCAAGATCCTCCTGCTGGACGAGGCC 1770
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
|||||
Db 1771 ACGTCAGCCTTTGGACACAGAAGTGAAGCAGTGGTTCAGSTGGCTCTGGATAAGGCCAGA 1830
QY 580 LysGlyArgThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
|||||
Db 1831 AAAGTCGACCAACCATTTGCTAGCTCATCTGTTGCTACGGTTCGTAATGCGCAGCTC 1890
QY 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
|||||
Db 1891 ATCGTGTGTTTCGATGATGGAGTCATTGTGGAGAAAGGAATCATGATGAGCTCATGAAA 1950
QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
|||||
Db 1951 CAGAAAGCATTTACTTCAAACTTGTCCAAATGCAGACAGCAGGAAATGAAATTTGAATTA 2010
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAlaLeuLeuMetSerProLysAsp 659
|||||
Db 2011 GAAATCGAGCTGATGAATCCAAAGTGAATTCATACCTTGGAAATGCTCTCATCATGAT 2070
QY 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
|||||
Db 2071 TCAGGATCCAGCTCTAATAAGAAAAAGATCCACTCGTAGGAGTGTCCGTGGATCACAAAGC 2130
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
|||||
Db 2131 CAACACAGAAAGCTTAGTACCAGAGGCTCTGGATGAAAGTATACCTCCAGTTTCCTTT 2190
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
|||||
Db 2191 TGGAGGATTATGAAGCTAAATTTAACTGAGTGGCCTTATTTGTGTGTGTATTTGT 2250
QY 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
|||||
Db 2251 GCCATTATAATGAGGCTGTCAACACGACATTTGCATTAATATATTTCAAAAGATTATAGG 2310
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
|||||
Db 2311 ATTTTACAGAAATGATGATGCGGAACAAACAGCAGAAATAGTACCTGTTTTCACATA 2370
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
|||||
Db 2371 TTGTTTCTAGCTCTGGAATGTTCTTTTATTACATTTTCTTCAGGGCTTCACATTT 2430
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
|||||

Db 2431 GGCAAAAGCTGAGAGATCCCTCAACAAGCGCTCCGATACATGTTTTCCGATCATGCTC 2490
QY 800 ArgGlnAspValSerTriPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
|||||
Db 2491 AGACAGGATGTGACCTGGTTGATGACCCCTAAAAACACCACTGGAGCATTTGACTACAGG 2550
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
|||||
Db 2551 CTCGCCAATGATGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTGCTATAATATACC 2610
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
|||||
Db 2611 CAGAAATAGCAAACTCTGGCAGAGGAATAATATATATCTTAACTATGTTGGCAACTG 2670
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
|||||
Db 2671 AACTGTTACTCTTAGCAATTTGACCCATCATTCGAATACAGGAGTTGTTGAATGAAA 2730
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluAlaGlyLysIleAla 899
|||||
Db 2731 ATGTTGTCTGGACAAGCACTGAAGATAGAAAAGAACTAGAAAGGTGCTGGGAAGATCGCT 2790
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
|||||
Db 2791 ACTGAACCAATAGAAAACITCCGAACCTGTTGTTCTTTGACTCAGGAGCAGAAAGTTGAA 2850
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
|||||
Db 2851 CATATGTATCATCAGAGTTTGGAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATC 2910
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
|||||
Db 2911 TTTGGAAATCAGTTTCTTCACGAGGCAATGATGATATTTTCTCTATGCTGGATGTTTC 2970
QY 960 ArgPheGlyAlaTyrIleuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
|||||
Db 2971 CGCTTTGGAGCCTACTTGGTGGCACATAGTCTCATCAGCTTTGAGGATGTTCTGTTAGTA 3030
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
|||||
Db 3031 TTTTTCAGCTGTTGCTTGTGTCATGCGCGTGGGCAAGTCAGTTCAATTTGCTCTCTGAC 3090
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
|||||
Db 3091 TATGCCAAAGCCAAAGATATCAGCGCCACATCATCATGATCATTTGAAAAAACCCCTTG 3150
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
|||||
Db 3151 ATTGACAGCTACAGCACAGAGGCCCTAAAGCCGAACACATTTGGAAGAAATGTCACATTT 3210
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
|||||
Db 3211 AATGAAGTTGTAATCAACTATCCACCCGACTGGACATCCCATGCTTCAGGGGTGAGC 3270
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
|||||
Db 3271 CTGGAAGTGAAGAGGCCAGCGCTGGCCCTGGTGGGCGACGATGGCTGTGGGAGAGC 3330
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
|||||
Db 3331 ACGGTGGTCCAGCTCTCGAGCGGTTCTATGACCCCTTGGCGGGGAAAGTGTGCTGTGAC 3390
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
|||||
Db 3391 GGCAAGAAATTAAGCACTGAATGTTCACTGGCTCCGAGCACACCTGGGCATCTGTGCC 3450
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
|||||
Db 3451 CAGGAGCCCATCCTGTTGACTGACGATTAGTGAGAACATTTGCCCTATGGAGACAAACAGC 3510
QY 1140 ArgValValSerHisGluGluIleValGlnAlaIleLysGluAlaAsnIleHisPhe 1159
|||||


```
Db 3511 CCGGTGGTGCACAGGAAGATCGTGAGCGCAGCCAGGAGGCAATATATACAGCCCTTC 3570
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
PS |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
Db 3571 ATCGAGTCACTGCCCTAATAATATAGCACCAGAGTAGGAGACAAGGAACTCAGCTCTCT 3630
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3631 GGTGCCCAAGAACACGCAATGTCATAGCTCGTCCCTTGTAGACAGCCCTCATATTTTG 3690
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3691 CTTTTGGATGAAGCCACATCAGCTCTGGATACAGAAAGTGAAGAGGTGTCCAAAGAGCC 3750
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3751 CTGGCAAAAGCCAGAGAAGCGGTACCTGCAATGTGATTGCTACCGCCTGTCCACCATC 3810
QY 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3811 CAGAAATGCAGACTTAATAGTGTGTTTCAGAAATGCAGAGTCAAGGAGCAGCGCACAT 3870
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3871 CAGCAGCTGCTGGCAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAGCA 3930
QY 1280 LysArg 1281
Db 3931 AAGCGC 3936

RESULT 9
AAF86128
ID AAF86128 standard; cDNA; 4195 BP.
XX
AC AAF86128;
XX
DT 25-JUN-2001 (first entry)
XX
DE Cynomolgus monkey P-glycoprotein cDNA variant 1.
XX
KW Cynomolgus monkey; P-glycoprotein; PGP; multidrug transporter; MDRI;
KW efflux pump; ss.
XX
OS Macaca fascicularis.
XX
FH Key Location/Qualifiers
FT CDS 100..3951
FT FT /*tag= a
FT FT /product= "PGP"
FT FT /note= "P-glycoprotein"
FT FT misc_feature 376..384
FT FT /*tag= b
FT FT /note= "Insertion of 9 nucleotides relative to PGP
FT FT allelic variant AAF86127"
XX
XX WO200123565-A1.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26592.
XX
XX 28-SEP-1999; 99US-0156921.
XX 12-OCT-1999; 99US-0158818.
XX
XX (GENT-) GENTEST CORP.
XX
XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
XX
XX WPI: 2001-316136/33.
XX P-PSDB; AAB81065.
XX
XX Novel isolated nucleic acid encoding cynomolgus monkey P-glycoprotein
XX (PGP) and homologous PGP polypeptides are useful for predicting
```

bioavailability of compound and increasing PGP transporter activity in cell

Example 1; Page 59-65; 84pp; English.

This invention relates to a polynucleotide sequence encoding a cynomolgus monkey P-glycoprotein (PGP), and an allelic variant of the PGP protein. PGP, also known as multidrug transporter, MDRI is a member of the ABC transporter superfamily. The enzyme serves as an efflux pump exporting small molecules across the cell membrane. The invention includes a cynomolgus monkey (Macaca fascicularis) PGP coding sequence and protein, and also that of an allelic variant. The PGP polynucleotide sequence is useful for increasing PGP transporter activity in a cell. Antisense sequences of the cDNA are useful for inhibiting PGP transport activity in a mammalian cell. They may also be used for increasing the bioavailability of a drug. The present sequence represents cDNA encoding cynomolgus monkey P-glycoprotein. This sequence contains a 9 nucleotide insert compared to the PGP allelic variant given in AAF86127.

Sequence 4195 BP; 1230 A; 801 C; 1039 G; 1125 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 4195
Score: 5859.50 Matches: 1165
Percent Similarity: 94.80% Conservative: 57
Best Local Similarity: 90.38% Mismatches: 52
Query Match: 90.52% Indels: 15
DB: 22 Gaps: 5

US-09-672-725C-27 (1-1281) x AAF86128 (1-4195)

QY 1 MetAspProGluGlyArgLysGlySerAlaGlu---LysAsnPheThrLysMetGly 19

Db 100 ATGGATCTTTGAAGGGACCGCAATGGAGGAGCAGAGAAGAACTTTTAACTGAAC 159

QY 20 LysLysSerLysLysLysGluLysLysLysProThrValSerThrPheAlaMet 39

Db 160 AATAAAGT---AAAAAGATAGAAGAAAGAAAGAACTGTCAGTGTATTTCATG 216

QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59

Db 217 TTTTCGCTATTCAAAATGGCTTGACAAAGTTGATATGCTGGTGGAACTTTGGCTGCATC 276

QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79

Db 277 ATCCATGGAGCTGGACTTCTCTCATGATGCTGGTGTGGAGACATGACGATACCTTT 336

QY 80 AlaAsnAlaGly-----IleSerArgAsnLysThrPheProVal 92

Db 337 GCAATGCAGGAAATTTAGGAGATTTAGGAGCTCTCTGTTTAAACAACACT----- 387

QY 93 IleIleAsnGluSerIleThrAsnAsnThrGlnHisPheIleAsnHisLeuGluGlu 112

Db 388 -----AATAGCAGTAATATATCACTGATACAGTCCCGCTCATGAAT---CTGGAGGAAGAT 438

QY 113 MetThrThrTyrAlaTyrTyrSerGlyIleGlyAlaGlyValLeuValAlaIleTyr 132

Db 439 ATGACCAGGTATGCCCTATTATTACAGTGGAAATGTGGCTGGGGTGTGCTGCTTAC 498

QY 133 IleGlnValSerPheTrpCysLeuAlaIleGlyArgGlnIleLeuLysIleArgLysGln 152

Db 499 ATTCAGTTTCAATTTGGTGCCTGGCAGCTGGAAGACAAATACACAAAATTAGAAAACAG 558

QY 153 PhePheHisAlaIleMetArgGlnGluIleGlyTyrPheAspValHisAspValGlyGlu 172

Db 559 TTTTTCATGCTATAATGCGACAGAGATAGGCTGTTGATGTGCACCATGTGTTGGGAG 618

QY 173 LeuAsnThrArgLeuThrAspValSerLysIleAsnGluGlyIleGlyAspLysIle 192

Db 619 CTTAACACCGGCTTACAGATGATGCTCCAAGATTAATGAAGGAATGTTGTCACAAAAT 678

QY 193 GlyMetPhePheGlnSerIleAlaThrPhePheThrGlyPheIleValGlyPheThrArg 212

|||||

Db 679 GGAATGCTCTTTCAGTCAATGCGCAACATTTTTCACCTGGCTTTATAGTAGGATTTTACACGT 738
QY 213 GlyTrpLysLeuThrLeuValIleLeuAlaIleSerProValLeuGlyLeuSerAlaAla 232
Db 739 GGTGGAACTAACTACCTTGTGATTTTGGCCATCAGTCCCTGTTCTTGGAGCTGCAGCTGCA 798
QY 233 IleTrpAlaLysIleLeuSerSerPheThrAspLysGluLeuLeuAlaIleAlaLysAla 252
Db 799 GTCTGGGCAAGATAGCTGCTTCAITACTGATAAAGAACTCTTAGCTTATGCAAAAGCT 858
QY 253 GlyAlaValAlaGluValLeuAlaAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 272
Db 859 GGAGCAGTAGCTGAAGAGCTCTGGCAGCAATTAGAAGCTGTAAGCTTGGGATAAG 918
QY 273 LysLysGluLeuGluArgTyrAsnLysAsnLeuGluAlaLysGlyIleGlyIleLys 292
Db 919 AGAAGAAGCTCGAAGGTACAAACAAATTTAGAGAAGCTTAAAGAAATTTGGGATAAG 978
QY 293 LysAlaIleThrAlaAsnIleSerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyr 312
Db 979 AAAGCTATTACAGCAATATTCTATAGTGTCTGCTTTCTCTGCTATCTATGCACTTAT 1038
QY 313 AlaLeuAlaPheTrpTyrGlyThrSerLeuValLeuSerSerGlyTyrThrIleGlyGln 332
Db 1039 GCTCTGGGCTCTGTATGGGACCACTTGGTCTCTCAAGGAATATCTATTGGACAA 1098
QY 333 ValLeuThrValPheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerPro 352
Db 1099 GTACTCACTGATCTCTTCTGTATTAATGGGGCTTTTAGTGTGGACAGCATCTCCA 1158
QY 353 SerIleGluAlaPheAlaAsnAlaArgGlyAlaAlaIleGlyIlePheLysIleIleAsp 372
Db 1159 AGCATTTGAAGCATTTGCAAAATGCAAGAGGAGCAGCTTTTGAATCTTCAAGATAATTGAT 1218
QY 373 AsnLysProSerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGly 392
Db 1219 AATAAGCCAAAGTATTGACAGCTATTTCGAAGAGTGGGCAACACAGATTAATTAAAGGA 1278
QY 393 AsnLeuGluPheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeu 412
Db 1279 AATTTGGAATTCAGAAATGTTCACTTTCAGTTACCATCTCGAAAGAAGTTAAATCTTG 1338
QY 413 LysGlyLeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGly 432
Db 1339 AAGGCCCTGAACTGGAAGGTGCAGAGTGGGACAGCGTGGCCCTGGTTGGAACACAGCGC 1398
QY 433 CysGlyLysSerThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMet 452
Db 1399 TGTGGGAAGACCAACCGTCCAGTGTGATGAGAGGCTTTATGACCCACACAGAGGGCATG 1458
QY 453 ValCysIleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThr 472
Db 1459 GTCAAGTGTGATGACAGGATATTAGGACCATTAACGTAGGTTTCTACGGGAATCATC 1518
QY 473 GlyValValSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyr 492
Db 1519 GGTGTGTGTGAGTACGAACCTGTATTGTTGCCACACGATAGCTGAAACATTCGCTAT 1578
QY 493 GlyArgGluAsnValThrMetAspGluIleGluLysAlaValLysGluAlaAsnAlaTyr 512
Db 1579 GTCGTGAAGATGTCACCTGGATGGATTTAGAAAGCTGTCAAGGAAGCCAAATGGCTAT 1638
QY 513 AspPheIleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGln 532
Db 1639 GACTTTATCATGAATTCCTCAGAAATTTGACACCTCGTTGGAGAGAGAGGGGCCAG 1698
QY 533 LeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLys 552
Db 1699 CTGAGTGTGTGGGACAGACAGAGATGCCATTCGACCTGCGCTGTGCTGCAACCCCAAG 1758
QY 553 IleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGln 572
Db 1759 ATCTCTGCTGGACGAGGGCCAGCTCAGCCTTGGACACAGAAAGTGAAGCAGTGGTTTCA 1818

QY 573 VallalaLeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSer 592
Db 1819 GTGGCTCTGGATAAGGCGCAGAAAGGTGGACACCATTTGTGATAGCTCATCGTTTGTCT 1878
QY 593 ThrValArgAsnAlaAspValIleAlaGlyPheAspGlyValIleValGluLysGly 612
Db 1879 ACGTTCGTAATGCCGACGTCTCGTGGTTTCGATGATGGAGTCATTTGTGGAGAAAGGA 1938
QY 613 AsnHisAspGluLeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThr 632
Db 1939 AATCATGATGAGCTCATGAAGAGAGAAAGGCATTTACTTCAAACTGTGCACAAATGCACAA 1998
QY 633 ArgGlyAsnGluIleGluLeuGluAsnAlaThrGlyGluSerLysSerGluSerAspAla 652
Db 1999 GCAGGAATGAATTTGAATTTAGAAATTCGACCTGATCAATCCAAAGTGAATTTGATACC 2058
QY 653 LeuGluMetSerProLysAspSerGlySerSerLeuIleLysArgArgSerThrArgArg 672
Db 2059 TTGGAATGCTTCCACATGATTCAGGATCCAGTCTAATAAGAAAGATCCACTCGTAGG 2118
QY 673 SerIleHisAlaProGlnGlyGlnAspArgLysLeuGlyThrLysGluAspLeuAsnGlu 692
Db 2119 AGTCTCGTGGATCACAAGGCCAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAA 2178
QY 693 AsnValProProValSerPheTrpArgIleLeuLysLeuAsnSerThrGluTrpProTyr 712
Db 2179 AGTATCTCCAGTTTCCCTTTTGGAGGATTTATGAAGTAAATTTAAGTGAAGTGGCCTAT 2238
QY 713 PheValValGlyIlePheCysAlaIleAlaAsnGlyGlyLeuGlnProAlaPheSerIle 732
Db 2239 TTTGTTGTTGTTATTTTGTGCCATTTAATAATGGAGTCTGCACACGAGATTTGGAGTA 2298
QY 733 IlePheSerArgIleIleGlyIlePheThrArgAspGluAspProGluThrLysArgGln 752
Db 2299 ATATTTTCAAGATATTAGGGATTTTACAAAGAAATGATGATCGCAAAACAAACGACAG 2358
QY 753 AsnSerAsnMetPheSerValLeuPheLeuValLeuGlyIleIleSerPheIleThrPhe 772
Db 2359 AATAGTAACTGTGTTTTCATATTGTTTCTAGTCTCTGGAAATTTGTTCTTTTATATATTT 2418
QY 773 PheLeuGlnGlyPheThrPheGlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyr 792
Db 2419 TTCTTTCAGGCTTTCACATTTGCAAAAGCTGGAGAGATCTCACCACGAGCTCCCATAC 2478
QY 793 MetValPheArgSerMetLeuArgGlnAspValSerTrpPheAspAspProLysAsnThr 812
Db 2479 ATGGTTTTCGATCCATGCTCAGACAGGATGTGAGCTGGTTTGTATGACCTTAAACACAC 2538
QY 813 ThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGly 832
Db 2539 ACTGGAGCATGTACTACGAGCTCGCCATGATGCTCAAGTTTAAAGGGGCTATAGGT 2598
QY 833 SerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSer 852
Db 2599 TCCAGGCTTGTCTAATAATCCAGATATAGCAAACTTTGGGACAGGAATAATTATATCC 2658
QY 853 LeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAlaIleValProIleIleAlaIle 872
Db 2659 TTAATCTATGTTGGCAACTGCACACTGTACTCTTAGCAATTTGTACCCATCATTCGAATA 2718
QY 873 AlaGlyValValGluMetLysMetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeu 892
Db 2719 GCAGGAGTGTGTGAATGAATATGTTGCTGGACAGCACTGAAGATTAAGAAAGAACTTA 2778
QY 893 GluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeu 912
Db 2779 GAAGGTGCTGGGAAGATCGCTACTGAAGCAATAGAAACTTCCGAACTGTTGTTCTTTTG 2838
QY 913 ThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsn 932
Db 2839 ACTCAGGACGAAAGTTTGAACATATGATGATCAGAGTTTGCAGGTATGACGTACCATACAGAAAC 2898

Qy	933	SerLeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyr	952
Db	2899	TCITTTGAGGAACACACATCTTTTGAATACAGCTTTTCTTCCACGACAGGCAATGATGAT	2958
Qy	953	PheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsn	972
Db	2959	TTTTTCTATGCTGGATGTTTCCGGTTTGAGGCTTACTTGGTGGCACATAGTCTCATGAGC	3018
Qy	973	PheGlnAspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGln	992
Db	3019	TTTGAGGATGTTCTGTAGTATTTTTCAGCTGTGTCTTTGGTGGCCATGGCGGTGGGCA	3078
Qy	993	ValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMet	1012
Db	3079	GTCAGTTTCATTTGCTCTGACTATGCCAAGCCAAAGTATCAGCAGGCCACATCATCATG	3138
Qy	1013	IleIleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThr	1032
Db	3139	ATCATTCGAAAAACCCCTTTTGATTGACAGCTACACGCACAGAGGCCCTAAAGCCGAACA	3198
Qy	1033	LeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIle	1052
Db	3199	TTGGAAGGAATGTCACATTTTAATGAAGTTGTATTCAACTATCCACCCGACTGGACATC	3258
Qy	1053	ProValLeuGlnGlyLeuSerLeuValLysLysGlyGlnThrLeuAlaLeuValGly	1072
Db	3259	CCAGTGTCTCAGGGCTGAGCTGGGAAGTGAAGAGGGCCAGACGCTGGCCCTGGTGGCC	3318
Qy	1073	SerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeu	1092
Db	3319	AGCAGTGGCTGTGGGAAGACAGCGGTGTCAGCTCCTGGAGCGTTCTATGACCCCTTG	3378
Qy	1093	AlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArg	1112
Db	3379	CGCGGGAAAGTGTGCTGTGACGCAAGAAATAAAGCAACTGAATGTTCAGTGGCTCCGA	3438
Qy	1113	AlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsn	1132
Db	3439	GCACACTGGGCATCGTGTCCAGAGAGCCCATCTCTGTGACTGCAGCATTAGTGAGAAC	3498
Qy	1133	IleAlaTyrGlyAspAsnSerArgValValSerHisGluGluIleValGlnAlaAlaLys	1152
Db	3499	ATTGCTCATGTGAGACAAACAGCCGGTGTGTACAGGAGAGATCGTGAGGCGACGCCAAG	3558
Qy	1153	GluAlaAsnIleHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGly	1172
Db	3559	GAGGCCAATATACAGCCTTCATCGAGTCACGTGCCTTAATAATATAGCACCCAGATAGGA	3618
Qy	1173	AspLysGlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeu	1192
Db	3619	GACAAAGGAATCAGCTCTCTGTTGCCAGAAACAGCATTTGCCATAGCTCGTGCCCTT	3678
Qy	1193	ValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAsnThrGluSer	1212
Db	3679	GTTAGACAGCCTCATATTTTGTCTTTTGGATGAAGCCACATCAGCTCTGGATACAGAAAGT	3738
Qy	1213	GluLysValValGlnGlnAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIle	1232
Db	3739	GAAAGGTTGTCCAAGAAGCCCTGGCAAGCCAGAGAGGCCGTACCTGCAATTGTGATT	3798
Qy	1233	AlaHisArgLeuSerThrIleGlnAsnAlaAspLeuIleValIlePheGlnAsnGlyLys	1252
Db	3799	GCTACCCGCCCTGTCCACCATCCAGATGACAGCTTAATAGTGGTGTTCAGAAATGGCAGA	3858
Qy	1253	ValLysGluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMet	1272
Db	3859	GTCAGGAGACAGGCACATCAGCAGCTGCTGGCACAGAAAGGCATCTATTTTTCATG	3918
Qy	1273	ValSerValGlnAlaGlyAlaLysArg	1281
Db	3919	GTCAGTGTCCAGGCTGGAGCAAGCGC	3945

AAZ49333	AAZ49333 standard; cDNA; 3860 BP.
ID	AAZ49333 standard; cDNA; 3860 BP.
XX	
AC	AAZ49333;
XX	
DT	14-MAR-2000 (first entry)
XX	
DE	Human G185V mutant multidrug resistance-1 (MDR-1) cDNA.
XX	
KW	Multidrug resistance; MDR-1; P-glycoprotein;
KW	transmembrane efflux pump; haematopoietic stem cell; transdu-
KW	bone marrow transplantation; chemotherapy; radiation therapy
KW	gene therapy; gene replacement; genetic defect; thalassemia
KW	Gaucher's disease; sickle cell anaemia; leukaemia; ex vivo
KW	cytokine; mutant; ds.
XX	
OS	Synthetic.
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..3843
FT	/*tag= a
FT	/product= "Human G185V mutant MDR-1 protein"
XX	
PN	W09961589-A2.
XX	
PD	02-DEC-1999.
XX	
PF	27-MAY-1999; 99WO-US11825.
XX	
PR	28-MAY-1998; 98US-0086988.
XX	
PA	(SUJD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX	
PI	Sorrentino B, Bunting K;
XX	
XX	WPI; 2000-072615/06.
DR	P-PSDB; RAY58187.
XX	
PT	Ex vivo expansion of hematopoietic stem cells transduced with
PT	sequence encoding human multidrug resistance-1, used for bone
PT	transplantation
XX	
PS	Example 1; Page 79-82; 113pp; English.
XX	
CC	This sequence represents cDNA encoding human G185V mutant
CC	resistance protein MDR-1, where the Gly residue at position
CC	of the wild-type protein (AAV58186) is replaced by Val. MDR-
CC	transmembrane efflux pump, responsible for the export of dru-
CC	cells, particularly cancer cells. The wild-type MDR-1 shows
CC	resistance to etoposide and decreased resistance to vinca al-
CC	compared with the G185V mutant. The invention relates to tra-
CC	haematopoietic stem cells with nucleic acid encoding an MDR
CC	and culturing the modified cells. The modified haematopoietic
CC	cells are useful in bone marrow transplantation (to reconsti-
CC	haematopoietic systems in patients who have undergone chemo-
CC	radiation therapy) and in ex vivo gene therapy of genetic de-
CC	cells derived from haematopoietic stem cells, e.g., thalassem-
CC	Gaucher's disease, sickle cell anaemia or leukaemia. The mod-
CC	cells can also be used to identify factors involved in regul-
CC	proliferation and differentiation in haematopoietic stem cel-
CC	Haematopoietic stem cells that express MDR-1 will be protecte-
CC	chemotherapeutic agents, so can be engrafted while the patie-
CC	undergoing chemotherapy. Expansion of (rare) haematopoietic
CC	provides sufficient cells to permit standard biochemical ana-
CC	Overexpression of MDR-1 allows cytokine-driven expansion of
CC	haematopoietic stem cells by at least 10-fold compared with
CC	of 4-fold in known procedures.
CC	

Alignment Scores:
Pred. No.:

Score:	5857.00	Matches:	1163
Percent Similarity:	95.01%	Conservative:	55
Best Local Similarity:	90.72%	Mismatches:	60
Query Match:	90.48%	Indels:	4
DB:	21	Gaps:	4

US-09-672-725c-27 (1-1281) x AAZ49333 (1-3860)

QY	1	MetAspProGluGlyGlyArgLysGlySerAla--GluLysAsnPhetRplysMetGly	19
DB	1	ATGGATCTGAAGGGCGCGCAATGGAGGACCAAGAAAGAACTTTTAAACTGAAC	60
QY	20	LysLysSerLysLysGlyLysLysGlyLysLysProThrValSerThrPheAlaMet	39
DB	61	AATAAAAGT---GAAAGAAATAAGAGGAAAGAAACCAACTGTCAGTGTATTTCATG	117
QY	40	PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle	59
DB	118	TTTCGCTATTCAAAATGGCTTGACAAGTGTATATGGTGGGAACTTTGGCTGCCATC	177
QY	60	IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe	79
DB	178	ATCCATGGGCTGGACTTCCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT	237
QY	80	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr	99
DB	238	GCAAATGCGAGGA---AATTTAGAAGATCTGATGTCAAACATCACAATAAGAGTGATC	294
QY	100	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrrAlaTyrTyr	119
DB	295	AATGATACAGGGTCTCTCATGAAT---CTGGAGGAAGACATGACAGATATGCTATAT	351
QY	120	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys	139
DB	352	TACAGTGAATGGTGGGCTGGTGGTCTGCTGCTTACATTGAGTTTCATTTTGGTGC	411
QY	140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159
DB	412	CTGGCAGCTGGAAGACAAATACAAAAATTAGAAAAACAGTTTTTTCATGCTATATGCGA	471
QY	160	GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179
DB	472	CAGAGATAGGCTGGTTGGTGGACGATGTTGGGAGGCTTAACCCGCACTTACAGAT	531
QY	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle	199
DB	532	GATGCTCTAAGATTATGAAGTTATGGTGACAAAATTGGAATGTTCTTTCATCAATG	591
QY	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219
DB	592	GCAACATTTTTCACGTGGTTTATAGTAGGATTTACACGTGGTTGGAAGCTAACCCTTGTG	651
QY	220	IleLeuAlaIleSerProValIleGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239
DB	652	ATTTTGGCCATCAGCTGCTGTTGGACTGTACGCTGCTGCTGGGAAAAGATACTATCT	711
QY	240	SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal	259
DB	712	TCATTTACTGATAAAGAACTCTTAGCGTATGCCAAAAGCTGGACGAGTAGCTGAAGAGTC	771
QY	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyr	279
DB	772	TTGGCAGCAATTAGAACTGTGATTCATTTGGAGGACAAAAGAAAGAACTTGAAGGTAC	831
QY	280	AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299
DB	832	ACAAAAATTTAGAAGAAAGCTAAAAGAAATTGGGATATAAGAAAGCTATTACAGCCAATAT	891
QY	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319
DB	892	TCATATAGTCTGCTTCTTCCTGCTATGATCTATGCTCTGCTGCTGCTGCTGCTGCTGCT	951
QY	320	ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer	339

DB	952	ACCACCTGGTCTCTCAGGGGAATATCTATTGGACAAGTACTCATCTATTCTTTTCT	1011
QY	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359
DB	1012	GTATTAATTTGGGCTTTTAGTTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAT	1071
QY	360	AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379
DB	1072	GCAAGAGAGCAGCTTATGAATCTTCAAGATATTGATAATAAGCCAGTATTGACAGC	1131
QY	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399
DB	1132	TATTCGAAAGAGTGGCCACAAACAGATATATTAAAGGAAATTTGGAATTCAGAATGTT	1191
QY	400	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419
DB	1192	CACPTCAGTTACCATCTCGAAAAAGATTAAAGATCTTGAAGGCGCTGAACCTGAAGGTG	1251
QY	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal	439
DB	1252	CAGAGTGGGACAGCGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGACACACAGTC	1311
QY	440	GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459
DB	1312	CAGCTGATGCAGAGGCTCTATGACCCACAGAGGGGATGTCAGTGTGTGATGGACAGGAT	1371
QY	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro	479
DB	1372	ATTAGGACCAATAATTAAGGTTTCTACGGGAAATCATTTGCTGTGTGATCAGGAACCT	1431
QY	480	ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499
DB	1432	GTATTTGTTGCCACCGATAGCTGAAACAACTTCGCTATGGCCGTGAAATGTCCACATG	1491
QY	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro	519
DB	1492	GATGAGATTGAGAAAGCTGTCAAGGAAGCCAAATGCCATGACATTTATCAAGAACTGCCT	1551
QY	520	AsnLysPheAspThrLeuValGlyArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539
DB	1552	CATAAAATTTGACACCCCTGGTTGGAGAGAGAGGGGCCAGTTGAGTGGTGGCAGAGCAG	1611
QY	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559
DB	1612	AGATGCCATTGCACGTGCCCTGGTTCCGACCCCAAGATCCTCCTGCTGATGAGGCC	1671
QY	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579
DB	1672	ACGTCAGCCTTGGACACAGAAAGCAGAGCAGTGGTTTCAGGTGGCTCTGGATGAAGCCAGA	1731
QY	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599
DB	1732	AAAGGTGCGACCACTTGTAGTATGTCATCGTTTCTACAGTTCGTATGCTGACGTC	1791
QY	600	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619
DB	1792	ATCGCTGGTTTCGATGATGAGTCATTGTGGAGAAAGAAATCATGATGAATCATGATAA	1851
QY	620	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu	639
DB	1852	GAGAAAGGCATTTACTTCAAACTGTCAACATGCACAGCAGCAGGAATGAAGTTGAATTA	1911
QY	640	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	659
DB	1912	GAAAAATGCAGCTGATGAATCCAAAAATGAATGATGCTTGGAAATGCTTCAATGAT	1971
QY	660	SerGlySerSerLeuIleLysArgSerThrArgSerIleHisAlaProGlnGly	679
DB	1972	TCAGATCCAGTCTAATAAGAAAAAGATCAACTCGTAGGAGTGTCCGTGGATGCACAGCC	2031
QY	680	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe	699

```
Db 2032 CAAGACAGAAAGCTTACTACCAAGAGCGCTCTGGATGAAAGTATACCTCCAGTTTCCTTT 2091
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2092 TGGAGGATTATGAGCTAAATTTAACTGAATGCCATTATTTGTTGGTGATTTTGT 2151
Qy 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2152 GCCATTATAANTGGAGCGCTGCACCCAGCATTTTGCAATAATATTTTCAAAAGATTATAGG 2211
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2212 GTTTTTCACAGAATTGATGATCTCTGAACAAACGACAGAAATAGTAACCTTGTTTTTCAC 2271
Qy 760 PhePheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2272 TTGTTTCTAGCCCTTGGAAATATTCTTTTATTACATTTTTCCTTCNAGGTTTTCACATTT 2331
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2332 GGCAAAAGCTGGAGAGATCCTCACCAAGCGCTCGGATACATGGTTTCCGATCCATGCTC 2391
Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2392 AGACAGGATGTGAGTTGGTTGATGACCCCTAAACACACCATGGAGCATTTGACTACCA 2451
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2452 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTGCTGTAATTACC 2511
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2512 CAGAATATAGCAAACTCTGGACAGGAATATATATCTCTCATCTATGTTGGCAACTA 2571
Qy 860 ThrLeuLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLys 879
Db 2572 ACACGTGTACTCTTAGCANAATTGTACCATCATTCATCATTAGCAGAGTTGTTGAATGAA 2631
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2632 ATGTTGCTCGACACAGCACTCAAGATAAGAAAGAACTAGAAAGTGTGCGGAAGATCGCT 2691
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2692 ACTGAAGCAANTAGAAAACCTCCGAACCGTTGTTCTTGACTCAGGACAGAAAGTTTGA 2751
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2752 CATATGATGTCAGAGTTTCGAGGTACCATACAGAAACTCTTTGAGGAAGCACACATC 2811
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2812 TTTGGAATTACATTTTCTTCACCAGCAATGATGATATTTTTCCTATGCTGGATGTTTC 2871
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 2872 CGGTTTGGAGCGCTACTTGGTGGCATAAACTCATAGCTTTGAGGATGTTCTGTTAGTA 2931
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 2932 TTTTCAGCTGTGCTTTGTTGGCCATGGCCGTGGGCAAGTCAGTTTCATTTGCTCCTGAC 2991
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 2992 TATGCCAAAGCCAAAATATACAGCAGCCACATCATCATGATCATTTGAAAAAACCCCTTG 3051
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3052 ATTGACAGCTACAGCAGGGAAGCCCTAATGCCGAACACATTGGAGGAATGTACATTT 3111
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3112 GGTGAAGTTGATTCAACTATCCACCCAGCGGACATCCCGAGTCTTCAGGGACTGAGC 3171
```

```
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3172 CTGAGAGTGAAAGAGGCCAGACGCTGGCTCTGGTGGCAGCAGTGGCTGTGGGAAGGC 3231
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3232 ACAGTGGTCCAGCTCCTGGAGCGGTCTACGACCCCTTGGCAGGAGAAAGTCTGCTGTAT 3291
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3292 GGCAAGAAATAAAGCGACTGAATCTTCAGTGGCTCCGAGCACACCTGGCATCGTGTC 3351
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3352 CAGGAGCCCATCCTCTTTGACTGCAGCATTTGCTGAGAACATTTGCCATATGGAGACAACAGC 3411
Qy 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3412 CGGGTGGTGTACAGAGAGATCGTGAGGCGCAGCAAGAGGCGCAACATACATGCTCTTC 3471
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3472 ATCGAGTCACTGCCTAAATAATATAGCACTAAAGTAGGAGACAAAGAACTCAGCTCTCT 3531
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3532 GGTGGCAGAAAACAGCATTCGCCATAGCTGTCGCCCTTTGTAGACAGCCTCATATTTTG 3591
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3592 CTTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAAGGTTGTCCCAAGAGCC 3651
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3652 CTGGACAAAGCCAGAGAGCGCACCTGCATTTGATTTGCTCACCGGCTGTCCACCATC 3711
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3712 CAGAATGCAGACTTAATAGTGGTGTTCAGATTCAGATTCAGAGAGTCAGAGGATCGCAGCAT 3771
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3772 CAGCAGCTGCTGGCAGACAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 3831
Qy 1280 LysArg 1281
Db 3832 AAGCGC 3837
RESULT 11
ABA94366
ID ABA94366 standard; DNA; 3860 BP.
XX
AC ABA94366;
XX
XX 26-MAR-2002 (first entry)
DT
XX
DE Human BCRP DNA related seq Id No. 3.
XX
KW Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;
KW cardiant; gene therapy; ds.
XX
OS Homo sapiens.
XX
FH Key
FT CDS 1..3843
FT /*tag= a
XX
PN WO200192877-A2.
XX
PD 06-DEC-2001.
XX
```

30-MAY-2001; 2001WO-US17459.
31-MAY-2000; 2000US-0584586.
29-MAY-2001; 2001US-0866866.
(SUUD-) ST JUDE CHILDREN'S RES HOSPITAL,
Sorrentino B, Schuetz J;
WPI; 2002-114368/15.
P-PSDB; ABB07267.
Identifying a stem cell, for treating e.g., muscular dystrophy,
myocardial infarction, Parkinson's disease, or neurodegenerative
disorders, comprises detecting the expression of an Atp transport
protein (BCRP) by a cell
Disclosure; Page 59-60; 87pp; English.
The invention provides a method of identifying and/or isolating a stem
cell that involves detecting the expression of an Atp transport protein
containing a conserved Atp-binding cassette (BCRP) by a cell in a sample
comprising stem cells. The isolated stem cells may be used in the
treatment of diseases such as muscular dystrophy, degenerative liver
disorder, myocardial infarction, Parkinson's disease, degenerative
disorders of the brain, and for tissue regeneration or replacement.
Haematopoietic cells can be used in bone marrow transplants (e.g., for
treatment of leukemia) and for ex vivo gene therapy for treating blood
diseases such as sickle cell anemia and thalassemia. The stem cells can
also be used as cell targets in gene therapy protocols. The present
sequence represents a sequence related to the BCRP for which no relevant
information has been provided in the specification.
Sequence 3860 BP; 1135 A; 746 C; 956 G; 1023 T; 0 other;

QY	140	LeuAlaLaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	150
DB	412	CTGGCAGCTGGAAACAAATACACAAATTAGAAAAACAGTTTTTTCATGCTATAATGCGA	471
QY	160	GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179
DB	472	CAGGAGATAGCGTGGTTTGATGTGCAGCATGTGGGAGCTTAACACCCGACTTACAGAT	531
QY	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle	199
DB	532	GATGCTCTAAGATTAAAGAATTATGGTGACAAAATTCGAATGTCCTTCAGTCAATG	591
QY	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219
DB	592	GCAACATTTTTCTACCTGGGTTTTATAGTAGGATTTACACGTGTTGGAAGCTAACCCCTTG	651
QY	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaIleTrpAlaLysIleLeuSer	239
DB	652	ATTTTGGCCATCAGTCCTGTTCTTGGACTGTCAGCTGCTGTCTGGCAAGATACATATCT	711
QY	240	SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluLeuVal	259
DB	712	TCATTTACTGATAAAGAACTCTTAGCGCTATGCCAAAAGCTGGACGACTACTGAAGAGTC	771
QY	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr	279
DB	772	TTGGCAGCAATTAGAACTGTGATTGCTATTTGGAGGCAAAAGAAAGAACTTCAAGAGTAC	831
QY	280	AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299
DB	832	AACAAAAATTTAGAAAGAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCAAATAT	891
QY	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319
DB	892	TCATATAGTGTCTTCTTCTCTGATCATCTATGCTCTGCCCTTCGGCTATGGG	951
QY	320	ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer	339
DB	952	ACCACCTGGTCCTCTCAGGGGAATATCTATTGGACAAGTACTCAGTGTATCTTTCT	1011
QY	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359
DB	1012	GTATTAAATGGGGCTTTTAGTGTGGACAGGCAATCTCCAAAGCATTTGAAGCATTTGCAAA	1071
QY	360	AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379
DB	1072	GCAAGAGGACAGCTTATGAATCTTCAGATAATTGATAATAAGCCAAAGTATTGACACG	1131
QY	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399
DB	1132	TATTGGAAGAGTGGGCACAAACAGATAATAATTAAAGGAAATTTGGAATTCAGAAATGTT	1191
QY	400	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419
DB	1192	CAC TTCAGTACCCATCTCGAAAGAAGTTAAGATCTTTGAAGGGCTCGAACCTGAAGGTG	1251
QY	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal	439
DB	1252	CAGAGTGGCAGACGGTGGCCCTGGTTGGAACAGTGGCTGTGGGAAGACGACCAACAGTC	1311
QY	440	GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459
DB	1312	CAGCTGATGCAGAGGCTCTATACCCACAGAGGGATGGTCAGTGTTGATGGACAGGAT	1371
QY	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGlnPro	479
DB	1372	ATTAGGACCATAAATTGAAGGTTCTACGGGAAATCATTTGGTGTGGTAGTCAGGAACCT	1431
QY	480	ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499
DB	1432	GTATTGTTGCCACACGATAGCTGAACAACATTCCTATGCCCCTGGAATAATGTCCACATG	1491
QY	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro	519

|||||
Db 1492 GATGAGATTGAGAAAGCTGTCAAGGAAGCCATGCCTATGACTTTATCATGAAATGCCT 1551
QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
:::|||||
Db 1552 CATAAATTTGACACCTGGTTGGAGAGAGAGGGGCCAGTTCAGTGGTGGCGAGAAGCAG 1611
QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
|||||
Db 1612 AGGATGCCATTGCAGTGCCTGGTTGCGAACCCCAAGATCCTCCTGCTGGATGAGGCC 1671
QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 1672 ACGTCAACCTTGGACACAGAAAGGAGCAGTGGTTCAGGTGGCTCGATRAAGGCCAGA 1731
QY 580 LysGlyArgThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 1732 AAAGGTCGGACCAACCAATTTGATAGCTCATCTGTTGTCTACAGTTCGTAATGCTGACGTC 1791
QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 1792 ATCCTCGTTTCGATGATGAGTCAATCTGTGGAGAAAGGAATCATGATGAACATGAAA 1851
QY 620 GluLysGlyIleIlePheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 1852 GAGAAAGGCATTTACTTCAAACTTGTCAAAATGCAGACAGCAGGAAATGAAGTTGAATTA 1911
QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 1912 GAAATGCAGTGTGATGAATCAAAAGTGAAATGATGCTTGGAAATGCTTCAAAATGAT 1971
QY 660 SerGlySerSerIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 1972 TCAGATCCAGCTCTAATAAGAAAGATCAACTCGTAGGAGTTCCTGGATCACAAGCC 2031
QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2032 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGTATACCTCCAGCTTCCTTT 2091
QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProIlePheValValGlyIlePheCys 719
Db 2092 TGGAGGATTATGAGCTAAATTTAACTGAAATGGCCTTATTTGTTGGTGTATTTGT 2151
QY 720 AlaIleLeuAsnGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2152 GCCATTATAATGGAGGCTGCAACCAAGCATTTCGAATAATATTTCAAAAGATTATAGG 2211
QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2212 GTTTTACAGAATTGTATGATCCTGGAACAAACGACAGAAATGTAACCTGTTTTCACTA 2271
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2272 TTGTTCTAGCCCTGGAAATTTCTTTTATTACATTTTCTTCAAGGTTTCACATTT 2331
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgIleMetValPheArgSerMetLeu 799
Db 2332 GCAAAAGCTGGAGAGATCCTCACCAGCGGCTCGGATACATGGTTTCCGATCCATGCTC 2391
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2392 AGACAGGATGTGAGTTGTTGATGACCTTAAAGGGCTATAGGTTCCAGGCTTGTGTAATTACC 2451
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2452 CTCGCCAATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGTGTAATTACC 2511
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleThrGlyTrpGlnLeu 859
Db 2512 CAGAAATATAGCAAAATCTTGGCAGGAAATATATATCTTCATCTATGTTGGCAACTA 2571
QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
|||||

Db 2572 ACACGTGTACTCTTAGCAAAATTGTACCCATCATTTGCAATAGCAGGAGTGTGTGAATGAAA 2631
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
|||||
Db 2632 ATGTTGCTGGCAGCACTGAAGATAAGAAAGAACTAGAAAGCTGTGGGAAGATCGCT 2691
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2692 ACTGAAGCAATAGAAAACCTCCGAACCGTTCTTCTTTGACTCAGGAGCAGAAAGTTTGA 2751
QY 920 TyrMetThrAlaGlnSerLeuGlnValProIleValProIleValProIleValProIle 939
Db 2752 CATATGTATGCTCAGAGTTTCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATC 2811
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetIleThrPheSerThrAlaGlyCysPhe 959
Db 2812 TTTGGAATACATTTTCTTCAACCCAGGCAATGATGATGATTTTCTCTATGCTGGATGTTT 2871
QY 960 ArgPheGlyAlaThrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 2872 CGGTTTGGAGCCTACTTGGTGGCACAATAACTCATAGAGCTTTGAGGATGTTCTGTAGTA 2931
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 2932 TTTTCAGCTGTGTTCTTGGTGCCATGCCGTGGGCAAGTCAGTTCATTTGCTCCTGAC 2991
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 2992 TATGCCAAGCCAAAATATCAGCAGCCACATCATCATGATCATTTGAAAAACCCCTTG 3051
QY 1020 IleAspSerThrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3052 ATTGCAGCTACAGCAGGAGGCTTATGCCGAACACATTGGAAGGAAATGTCACATTT 3111
QY 1040 AsnGluValValPheAsnThrProIleProIleProIleProIleProIleProIle 1059
Db 3112 GGTGAAGTTGATTAACATATATCCACCGCAGGACATCCAGTCTTCCAGGAGTGAAGC 3171
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3172 CTGGAGGTGAAGAAGGGCCAGACGCTGGCTGTGGGCGAGCAGTGGCTGTGGGAAGAGC 3231
QY 1080 ThrValValGlnLeuLeuGluArgPheThrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3232 ACAGTGGTCCAGCTCCTGGAGCGGTTCTACAGCCCTTGGCAGGAAAGTGTGCTGTGAT 3291
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSer 1119
Db 3292 GGCAAAAGAAATAAAGCGACTGAATGTTCAAGTGGCTCCGAGACACCTGGGCACTCGTCC 3351
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaIleThrGlyAspAsnSer 1139
Db 3352 CAGGAGCCCATCCTGTTTACTGACGATGCTGAGAACATTCCTTATGAGACACACAGC 3411
QY 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3412 CGGCTGGTGTCCAGCAGAGATCGTGGGCGCAGCAAGAGGCGCAACATACATGCTCTC 3471
QY 1160 IleGluThrLeuProGluLysThrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3472 ATCGAGTCACTGCTGCTAATAATATAGCACTAAAGTAGGAGACAAAGGAACTCAGCTCTC 3531
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3532 GGTGGCCAGAAACACGCAATTCCTGAGTCCGCTGCTGAGACAGCCTCATATTTTG 3591
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3592 CTTTGGATGAAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGAGGTTGTCCAGAAGGCC 3651
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
|||||
Db 3652 CTGGCAAAAGCCAGAGAAAGGCCGACCTGATTTGCTGCTACCCGCTGTCCACCATC 3711

QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
|||||
Db 3712 CAGAAAGCAGACTTATAGTGGTTCAGATGGCAGAGTCAAGGAGCATGGCAGCAT 3771

QY 1260 GlnGlnLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
|||||
Db 3772 CAGCAGTGTGGCAGACAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 3831

QY 1280 LysArg 1281
|||||
Db 3832 AAGCGC 3837

RESULT 12

AAH57442

ID AAH57442 standard; cDNA; 4349 BP.

XX

AC AAH57442;

XX

DT 10-SEP-2001 (first entry)

XX

DE Human Intestine cell specific cDNA sequence SEQ ID NO:282.

XX

KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX

OS Homo sapiens.

XX

PN WO200132927-A2.

XX

PD 10-MAY-2001.

XX

PF 02-NOV-2000; 2000WO-US30396.

XX

PR 04-NOV-1999; 99US-0163508.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Sornasse T, Seilhamer JJ, Watson GA;

XX

DR WPI; 2001-291057/30.

XX

PT New cell and tissue specific polynucleotides useful for diagnosis,
PT prognosis or monitoring of treatments for disorders where the gene is
PT associated with a cancer, immunopathology or neuropathology -

XX

PS Claim 1; Page 207-208; 327pp; English.

XX

CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC sequences (I). (I) can have cytostatic, immunomodulatory and
CC neuroprotective activities, and can be used in gene therapy. (I) and
CC proteins (II) encoded by then are used in high throughput screening
CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC their fragments, immunoglobulins, inhibitors, drug compounds and
CC pharmaceutical agents. Expression of (I) in a sample indicates the
CC differentiation of embryonic stem cells into a tissue selected from
CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC tissues. (I) and (II) are used to produce an expression profile that
CC defines a metabolic or developmental process, treatment, condition,
CC disease or disorder. The gene profile can be used for diagnosis,
CC prognosis or monitoring of treatments and for investigating a
CC predisposition to a disorder where the gene is associated with a
CC cancer, immunopathology or neuropathology.

SQ Sequence 4349 BP; 1304 A; 816 C; 1048 G; 1181 T; 0 other;

Alignment Scores:

Pred. No.:

Score:

Percent similarity:

Length:

Matches:

Conservative:

Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 22 Gaps: 4

US-09-672-725C-27 (1-1281) x AAH57442 (1-4349)

QY 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
|||||
Db 126 ATGGATCTTGAAGGGACCGCAATGAGGAGGAGCAAGAAAGAACTTTTAAACTGAAC 185

QY 20 LysLysSerLysLysGluLysLysGluLysProThrValSerThrPheAlaMet 39
|||||
Db 186 AATAAAAGT---GAAAAGATAAGAGAAAGAAACCACTGTCAGTGTATTTCAATG 242

QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAla 59
|||||
Db 243 TTTCGCTATTCAAATTTGGCTTGCACAAAGTTGATATGTTGGTGGGAACCTTTGGCTGCCATC 302

QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
|||||
Db 303 ATCCATGGGGCTGGACTTCCTCTCATGATCTGGTGTGGAGAAATGACAGATATCTTT 362

QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
|||||
Db 363 GCAATGCGAGCA---AATTTAGAGATCTGATGTCAACATCACTATAGAGTGATATC 419

QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
|||||
Db 420 AATGATACAGGGTCTTTCATGAAT---CTGAGGAGGAGACATGACAGGTATGCCTAATAT 476

QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
|||||
Db 477 TACAGTGAATTTGGTGGGGTGTGCTGCTTACATTTCAAGTTTCATTTGGTGC 536

QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
|||||
Db 537 CTGGCAGCTGGAACACAAATACACAAATTAGAAAAACAGTTTTCATGCTATANTCGGA 596

QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
|||||
Db 597 CAGGAGATAGCTGGTTTGTATGTGCACGATGTTGGGAGCTTACACCCGACTTACAGAT 656

QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle 199
|||||
Db 657 GATGCTCTAAGATTAAGAGTTATTTGGTGCACAAATTTGAAATGTTCTTTCAGTCAATG 716

QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
|||||
Db 717 GCAACATTTTTCACCTGGGTTTATAGTAGGATTTACACGTGGTTGGAGGCTAACCCCTGTG 776

QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
|||||
Db 777 ATTTTGGCCATCAGTCTCTGTTTGGACTGCTGCTGCTGCGGCAAGATACTATCT 836

QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
|||||
Db 837 TCATTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 896

QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
|||||
Db 897 TTGGCAGCAATTAGAACTGTGATTTGGAGGACAAAGAAAGAACTTGAAGAGTAC 956

QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
|||||
Db 957 AACAAAAATTTAGAAAGCTAAAGAAATTTGGGATATAAGAAAGCTATTACAGCAATATT 1016

QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
|||||
Db 1017 TCTATAGGTGCTGCTTCCTGCTGATCTATGCACTTATGCTGCGCTTCTGGTAGGG 1076

QY 320 ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer 339
|||||
Db 1077 ACCACCTTGTCTCTCAGGGGAATATTCTATTGGACAGTACTACTGTTCTTTCTTCT 1136

QY	340	VallLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359	QY	700	TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	719
DB	1137	GTATTAAATGGGGCTTTAGTGTGGACAGCATCTCCAAGCATTTGAGCAAT	1196	DB	2217	TGGAGGATTAATGAAGCTAAATTTAACTGAATGGCCCTATTTTGTGTGGTGTATTTTGT	2276
QY	360	AlaArgGlyAlaAlaTyrGluIlePheLysIleLeuAspAsnLysProSerIleAspSer	379	QY	720	AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	739
DB	1197	GCAAGAGGAGCAGCTTATGAATCTTCAAGATAATTTGATAATAAGCCAAGTATTGACAGC	1256	DB	2277	GCCATTATAATTTGGAGGCTGCAACACGATTTGCAATAATAATTTTCAAGATATATAGG	2336
QY	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399	QY	740	IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal	759
DB	1257	TATTTCGAAGAGTGGCCACAAACAGATAATTAAGGGAAATTTGGAATTCAGAAATGTT	1316	DB	2337	GTTTTACAGAAGATTGATGATCCTGAAACAAAAACACAGAAATAGTAACCTGTTTTCACTA	2396
QY	400	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419	QY	760	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	779
DB	1317	CAC TTCAGTTACCCATCTCGAAGAAAGAGTTAAGATCTTGAAGGCCCTGAACCTGAAGGTG	1376	DB	2397	TTGTTTCTAGCCCTTGGAAATTTCTTTTATTACATTTTTCCTTCAGGGTTTTCACATTT	2456
QY	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal	439	QY	780	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	799
DB	1377	CAGAGTGGCCAGACGGTGGCCCTGGTTGGAACACAGTGCCTGTGGGAAGACACAACAGTC	1436	DB	2457	GGCAAGCTGGAGAGATCCTCACAAGGGCTCCGATACATGGTTTCCGATCCATGCTC	2516
QY	440	GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459	QY	800	ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg	819
DB	1437	CAGCTGATGCGAGAGGCTCTATGACCCACAGAGGGGATGGTCAGTGTGATGGACAGGAT	1496	DB	2517	AGACAGGATGTGAGTTGGTTTGTGATGACCCCTAAAAACACACCTGGAGCATTTGACTACCAGG	2576
QY	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro	479	QY	820	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	839
DB	1497	ATTAGGACCATAAATGTAAGGTTTCTACGGGAATCATTTGGTGGTGAGTCAGGAACCT	1556	DB	2577	CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTGCTGTAATACC	2636
QY	480	ValLeuPheAlaThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499	QY	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	859
DB	1557	GTATTGTTGGCCACGACGATAGCTGAAACATTCGCTATGCGCCGTGAAATGTCACCATG	1616	DB	2637	CAGAATATAGCAAACTTGGGACAGGAATAATATATATCTTCATCTATGTTGGCACTA	2696
QY	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro	519	QY	860	ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	879
DB	1617	GATGAGATTGAGAAGCTGTCAAGGAAGCCATGCCCTATGACTTATCATGAACACTGCCCT	1676	DB	2697	ACACTCTTACTCTTAGCAATTTGATCCCATCTTGAATAGCAAGAGGTTGTTGAAATGAA	2756
QY	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539	QY	880	MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla	899
DB	1677	CATAAATTTGACACCTGGTTGGAGAGAGAGGGGCCAGTTGAGTGGTGCGCAGAACAG	1736	DB	2757	ATGTTGCTGGACACGACTGAAAGATAAGAAAGAACTAGAAGGTGCTGGGGAAGATCGCT	2816
QY	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559	QY	900	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu	919
DB	1737	AGGATCGCCATTTGACGTGCGCCTGGTTCGCAACCCCAAGATCCCTGCTGGATGAGGCC	1796	DB	2817	ACTGAAGCAATAGAAACTTCCGAACCGTTGTTCTTCTTCACTCAGGAGCAGAAGTTGAA	2876
QY	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579	QY	920	TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle	939
DB	1797	ACGTCAGCCTTGGACACAGAAGCGAAGCAGCTGGTTCAAGGTGGCTCTGGATAAGGCCAGA	1856	DB	2877	CATATCTATGCTCAGAGTTTGACAGTACCACAGCAACTCTTTGAGGAAGACACACATC	2936
QY	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599	QY	940	PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe	959
DB	1857	AAAGTCGGACCACCATTTGTATAGCTCATCGTTTGTCTACAGTTTCGTAATGCTGACGTC	1916	DB	2937	TTTGGAAATTACATTTTCCCTCACCAGGCAATGATGATATTTTCTATGCTGGATGTTTC	2996
QY	600	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619	QY	960	ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal	979
DB	1917	ATCGCTGCTTTCGATGATGAGTCAATTTGTGAGAAAGAAATCATGATGAACCTCATGAAA	1976	DB	2997	CGGTTTGGAGCCTTACTTGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA	3056
QY	620	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu	639	QY	980	PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValValSerSerPheAlaProasp	999
DB	1977	GAGAAAGGCATTTACTTCAAACTTGTACATGCAGACAGCAGGAATGAAGTTGAATTA	2036	DB	3057	TTTTTCAGCTGTGTTCTTGTGTCATGGCCGCGGCAAGTCCAGTTCATTTGCTCCTGAC	3116
QY	640	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	659	QY	1000	TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu	1019
DB	2037	GAATAATGCAGCTGATGAATCCAAAAGTGAATTTGATGCTTGGAAATGTCTTCAAAATGAT	2096	DB	3117	TATGCCAAAGCCAAATATACAGCGCCACATCATCATCATCATTTGAAAAAACCCTTTG	3176
QY	660	SerGlySerSerLeuIleLysArgArgSerThrArgSerIleHisAlaProGlnGly	679	QY	1020	IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe	1039
DB	2097	TCAAGATCCAGTCTAATGAAGAAAGATCAACTCTGAGGATGCTCCGCGGATCAACGCC	2156	DB	3177	ATTGACAGCTTACAGCACGGAAGGCCCTAATGCCACACATTTGGAGGAAATGTCTACATTT	3236
QY	680	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	699	QY	1040	AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer	1059
DB	2157	CAAGACAGAAAGCTTAGTACCAAAAGAGGCTCTCGGATGAAAGTATACCTCCAGTTTCTT	2216	DB	3237	GGTCAAGTGTATTTAACTATCCACCCGACCGACATCCAGTGCTTCAGGACTGACG	3296
				QY	1060	LeuGluValLysLysGlyGlnThrLeuAlaLeuValIcLysSerSerGlyCysGlyLysSer	1079


```
Db 3297 CTGGAGTGGAAGAGGCGCAGACGCTGCTCTGCTGGCGACGACTGCTGGGGAAGC 3356
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3357 ACAGTGGTCCAGCTCCCTGGAGCGGTCTACGACCCCTGGCAGGGAAGTGGCTGTGAT 3416
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3417 GGCAAGAAATAAGCGACTGAATGTTCACTGGCTCGAGCACACCTGGCGATCGTGTC 3476
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3477 CAGGAGCCATCCCTGTTGACTGCAGCATTCCTGAGAACAATTCCTATGGAGACAACAGC 3536
QY 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3537 CGGTGGTGTGCACAGGAAGAGATCGTGAGGCGACCAAGAGGCGCAACATACATGCTTC 3596
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3597 ATCGAGTCACCTGCTAATAATATAGCACTAAGTAGAGACAAGGAAGCACTAGCTCTCT 3656
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3657 GGTGGCCAGAAACAACGATTCGCATAGCTGCTGCCCTTGTAGACAGCCTCATATTTG 3716
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3717 CTTTTGGATGAAGCCAGCTCAGCTCTGATACAGAAAGTGAAGAGTTGCCAAGAAGCC 3776
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3777 CTGGACAAGCCAGAGAGCGCCGCTGCTGATGCTGCTCACCCTGCTCCACCATC 3836
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3837 CAGAATCGAGACTTAATAGTGGTCTTTCAGATGGCAGAGTCAAGGAGCATGGCACGCAT 3896
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3897 CAGCAGCTGTGGCAGAGAAGGCATCTATTTTCAATGGTCACTGCTCCAGGCTGGAACA 3956
QY 1280 LysArg 1281
Db 3957 AAGCGC 3962
RESULT 13
ID AAZ94738
AC AAZ94738 standard; cdna; 4646 BP.
XX
AC AAZ94738;
XX
XX 01-AUG-2000 (first entry)
DT
DE Human ATP binding cassette ABCB1 (MDR1) cdna.
XX
KW ABCB1; ATP binding cassette; human; cholesterol; lipid disorder;
KW atherosclerosis; lipid disorder; dyslipidemia; psoriasis;
KW lupus erythematosus; diagnosis; gene therapy; MDR1;
KW multidrug resistance; chromosome 7q21; ss.
XX
OS Homo sapiens.
XX
XX WO200018912-A2.
PN
XX
PD 06-APR-2000.
XX
PF 21-SEP-1999; 99WO-EP06991.
XX
PR 25-SEP-1998; 98US-0101706.
XX
PA (FARB ) BAYER AG.
```

```
XX Schmitz G, Klucken J;
PI WPI; 2000-293151/25.
XX
PT Adenosine triphosphate binding proteins useful for identifying agents
PT for treating atherosclerosis and other inflammatory disorders -
XX
PS Claim 9; Page 110-112; 154pp; English.
XX
CC The present sequence is that of human ATP binding cassette
CC subfamily B protein ABCB1 cdna. The cdna was identified using a
CC differential display method in which monocytes from peripheral
CC blood were subjected to macrophage differentiation and cholesterol
CC loading with acetylated low density lipoproteins and subsequent
CC degrading with high density lipoprotein (HDL3) to identify
CC cholesterol sensitive genes. The gene maps to chromosome 7q21
CC and is also termed MDR1 (multidrug resistance). The invention
CC provides cholesterol-sensitive ABC genes (see AAZ94734-63). These
CC genes, and polypeptides encoded by them, can be used for diagnostic
CC and therapeutic applications, and for biochemical or cell-based
CC assays to screen for pharmacologically active modulator compounds
CC useful for the treatment of lipid disorders, atherosclerosis or
CC other inflammatory diseases such as psoriasis and lupus
CC erythematosus.
XX
SQ Sequence 4646 BP; 1371 A; 892 C; 1129 G; 1254 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 4646
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 21 Gaps: 4

US-09-672-725c-27 (1-1281) x AAZ94738 (1-4646)
QY 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 425 ATGGATCTTTGAAGGGGACCGCAATGGAGGAGCAAGAAAGAACTTTTAACTGAAC 484
QY 20 LysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 485 AATAAAGT---GAAAGACATAAGAGAGAAAGAAACCAACTGTCAGTGTATTTCAATG 541
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
Db 542 TTTTCGCTATTCAAATTTGGCTTGACAAGTTGTATATGGTGGGAACTTTGGCTGCCATC 601
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 602 ATCCATGGGGCTGGACTTCCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 661
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 662 GCAATGCAAGG---AATTTAGAAGATCTGATGCTCAACATCATCAATAGAAGTATATC 718
QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 119
Db 719 AATGATACAGGGTCTTTCATGAAT---CTGGAGGAGAGACATGACCAGGTATGCTTAT 775
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaIleTyrIleGlnValSerPheTrpCys 139
Db 776 TACAGTGAATTTGGTCTGGGTGCTGCTGTGCTTACATTCAGGTTTCATTTTGGTGC 835
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 836 CTGGCAGCTGGAAGACAAATACACAAAATAGAAAACAGTTTTTTCATGCTATAATGCGA 895
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 896 CAGGAGATAGGCTGTTTGTATGTGCACCATGTTGGGAGCTTAACACCCGACTTACAGAT 955
```


|||||
Db 3116 ACTGAAGCAATAGAAAACCTCCGAACCGTGTCTTCTTGACTCAGGAGCAGAAGTTTGAA 3175
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGATGCTCAGAGTTTCAGGTACCATACAGAAACTCTTTGAGGAAGCACACATC 3235
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTGGAATTACATTTCTTCCACCCAGCAATGATGATTTTCTCTATGCTGGATGTTTC 3295
QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 3296 CGGTTTGGAGCCTATTGTTGGCCATGAGCAATAATCATGAGCTTTGAGGATGTTCTGTAGTA 3355
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3356 TTTTCAGCTGTGCTTTGTTGGCCATGAGCGTGGGCAAGTCAATTCATTTGCTCCTGAC 3415
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAAGCAAAATATATCAGCAGCCCATCATCATCATGATCATTTGAAAAACCCCTTG 3475
QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTGACAGCTACAGCAGCGGAAGCGCTTAATGCCGAACACATTTGGAAGAAATGTCACATT 3535
QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnLysSer 1059
Db 3536 GGTGAAGTTGTATTAACATATCCACCAGCGGACATCCCATGCTTCAGGGACTGAGC 3595
QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3596 CTGGAGGTGAAGAGCGCCAGACGCTGCTGTGGTGGGAGAGTGGCTGGGAGAGC 3635
QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3636 ACAGTGGTTCAGCTCTGAGGCGGTCTACGACCCCTTTGGCAGGAAAGTGTCTGTGAT 3715
QY 1100 GlyLysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyLysValSer 1119
Db 3716 GGCAGAAATAAAGCGACTGAATGTCAGTGGCTCCAGACACACCTGGGCATCGTGC 3775
QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 CAGGAGCCCATCTCTTTGACTGCAGCATTTGCTGAGAACATTTGCTATGGAGACACAGC 3835
QY 1140 ArgValValSerHisGluGluIleValGlnAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3836 CGGGTGGTGCAGAGAGATCGTGGGCGCAGCAAGAGGCGCAACATACATGCTCTC 3895
QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCGAGTCACTGCCTTAATAATATAGCACTAAAGTAGGAGACAAAGAACTCACTCTCT 3955
QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GTGGCCAGAAACACGATTTGCCATAGCTGCTGCCCTTTGTAGACAGCCTCATATTTG 4015
QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGGATGAAGCCACGCTCAGCTCTGGATACAGAAAGTGAAGAGGTGTCCAAGAGCC 4075
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 GTGGACAAAGCAGAGAGCGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4135
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGTCAAGGAGCATGTCAGCAT 4195
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
|||||

Db 4196 CAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGCTCAGTGTCCAGGCTGGAACA 4255
QY 1280 LysArg 1281
|||||
Db 4256 AAGCGC 4261
RESULT 14
AAT13394
ID AAT13394 standard; DNA; 6505 BP.
XX
AC AAT13394;
XX
DT 24-JUN-1996 (first entry)
XX
DE Hybrid vector pSF-MDR.
XX
KW Hybrid; vector; gene transfer; gene therapy; haematopoietic stem cell;
KW retroviral; murine embryonic stem cell virus; MESV;
KW Moloney murine sarcoma virus; (MoMuSV);
KW Friend murine leukaemia virus; F-MuLV; ds.
XX
OS Synthetic.
XX
PN DEL19503952-Al.
XX
PD 14-MAR-1996.
XX
PF 07-FEB-1995; 95DE-1003952.
XX
PR 08-SEP-1994; 94DE-1431973.
XX
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
PI Baum C, Ostertag W, Stocking-harbers C, Stockingharbers C;
XX
DR WPI; 1996-152306/16.
XX
PT Hybrid retroviral vectors - for gene transfer into haematopoietic
PT stem cells
XX
PS Disclosure; Page 25-29; 42pp; German.
XX
CC New hybrid vectors comprise (1) a leader region including the U5
CC region and trna primer binding site of murine embryonic stem cell
CC virus (MESV) or Moloney murine sarcoma virus (MoMuSV), and (2) a 3'-
CC LTR including the U3 and R regions of a Friend murine leukaemia
CC virus (F-MuLV). The vectors are useful for ex-vivo or in-vivo gene
CC therapy. High levels of gene transfer can be achieved in
CC haematopoietic stem cells and their myeloid (non-lymphatic) progeny.
CC pSF1, pSF2, pSF3 and pMW1 (sequences given in AAT13390-T13393) are
CC examples of such vectors.
CC Vector pSF-MDR (sequence given in AAT13394) is based on the
CC MESV vector R224.
XX
SQ Sequence 6505 BP; 1719 A; 1506 C; 1606 G; 1674 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 6505
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 17 Gaps: 4

US-09-672-725C-27 (1-1281) x AAT13394 (1-6505)
QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1817 ATGATCTTGAAGGGGCGCCGATGGAGGAGCAAGAGAACTTTTAACTGAAC 1876
QY 20 LysLysSerLysLysLysGluLysLysGluLysLysLysLysLysLysLysLysLysLysLys 39
Db 1877 AATAAAGT---GAAAGAGTAAGAGGAAAGAAACCACTGCTCAGTGTATTTTCAATG 1933

Db 4088 TTTGTTCTAGCCCTTGGAAATATTTCTTTTATACATTTTCCCTCAGGGTTTCACATTT 4147
 QY 780 GYLysAlaGlyGluLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
 Db 4148 GCGAAAGCTGGAGAGATCTCTCACCAGCGGCTCCGATACATGTTTCCGATCATCCTC 4207
 QY 800 ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
 Db 4208 AGACAGGATGAGCTTGGTTTGATGACCCCTAATAAACACCACTGGAGCATTTGACTACCAGG 4267
 QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
 Db 4268 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGTTCACGGCTTGCTGTAATTACC 4327
 QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
 Db 4328 CAGAAATAGCAATCTTGGGACAGGAATAATTATCTCTCATCTATGTTGGCAACTA 4387
 QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
 Db 4388 ACACTGTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGGATTTGTTGAATGAAA 4447
 QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluAlaGlyLysIleAla 899
 Db 4448 ATGTGTCTGGACAAGCACTGAAGATAAGAAAGAACTAGAGGTGCTGGGAAGATCGCT 4507
 QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
 Db 4508 ACTGAGCAATAGAAAACCTTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAAAGTTTGA 4567
 QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
 Db 4568 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAACTCTTTGAGGAAGCACATC 4627
 QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
 Db 4628 TTTGGAATATACATTTCTTCCCTTCACCCAGGCAATGATGATTTTCTCTATGCTGATGTTTC 4687
 QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
 Db 4688 CGGTTGGAGCCCTACTTGGTGGCACAATAAATCATGAGCTTTGAGGATGTTCTGTAGTA 4747
 QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
 Db 4748 TTTTCAGCTGTTGCTTGTGTGGCATGGCGTGGGCAAGTCACTGATTCATTTGCTCTGAC 4807
 QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
 Db 4808 TATGCCAAAGCCAAATATACAGAGCCCATCATCATCATCATCATCATCATCATCATCATCAT 4867
 QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
 Db 4868 ATTGACAGCTACAGCAGGAGCCCTAATCCGCAACACATTTGGAAGAAATGTCACATTT 4927
 QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
 Db 4928 GGTGAAGTTGTTATTCAACTATCCACCCGACGACATCCAGTCTTCAGGGACTGAGC 4987
 QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
 Db 4988 CTGGAGTGAAGAAGGCCAGAGCTGGCTGTGGTGGGAGCAGTGGCTGTGGGAAGAGC 5047
 QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
 Db 5048 ACAGTGGTCCAGCTCTCGAGCGGTCTACGACCCCTTGGCAGGGAAGAGTGGCTGTGAT 5107
 QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
 Db 5108 GCGAAAGAAATAAAGGACTGAATGTTCACTGGCTCCGAGCACACCTGGGCATCGTGCC 5167
 QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139

Db 5168 CAGGAGCCCATCTCTGTTGACTGCAGCATTTGCTGAGAAATTCCTTATCGAGACAACAGC 5227
 QY 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
 Db 5228 CGGTTGGTGTACAGGAAGAGATCGTGAGGGCAGCAAGAGGCAACATACATGCTTCTC 5287
 QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
 Db 5288 ATCGAGTCACTGCCTTAATAATATAGCACTAAAGTAGTAGGACAAAGAACTCAGCTCTCT 5347
 QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
 Db 5348 GGTGGCCAGAAACACGATTCGATAGCTCGTGGCCCTTGTAGACAGGCTCATATTTTG 5407
 QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
 Db 5408 CTTTTCGATGAAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAGCC 5467
 QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
 Db 5468 CTGGCAAAAGCCAGAGAGCCGACCTGCTGATTTGCTACCGCCTGTCCACCATC 5527
 QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
 Db 5528 CAGATGCAGACTTAATAGTGTGTTTCAAGATGGCAGAGTCAAGAGCATGGCAGCAT 5587
 QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
 Db 5588 CAGCAGCTCTGCGCAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAACA 5647
 QY 1280 LysArg 1281
 Db 5648 AAGCGC 5653
 RESULT 15
 AAZ24041
 ID AAZ24041 standard; cDNA; 8630 BP.
 XX
 AC AAZ24041;
 XX
 DT 04-FEB-2000 (first entry)
 XX
 DE Retroviral M4 mdr-1 cDNA.
 XX
 KW Retroviral vector; gag gene; gene therapy; chemotherapeutic agent;
 KW hematopoietic stem cell transformation; mdr-1; ss.
 XX
 OS Retrovirus.
 XX
 PN EP955374-A2.
 XX
 PD 10-NOV-1999.
 XX
 PF 07-MAY-1999; 99BP-0250151.
 XX
 PR 08-MAY-1998; 98DE-1022115.
 XX
 PA (PETT-) PETTE INST HEINRICH.
 XX
 PI Ostertag W, Baum C, Hildinger M;
 DR WPI; 2000-001087/01.
 XX
 PT New retroviral vector containing minimal or no gag gene sequence, for
 use e.g. in gene therapy or cloning -
 XX
 PS Disclosure; Page 16-18; 35pp; German.
 XX
 CC This invention describes a novel retroviral vector (RV) containing a gag
 CC gene fragment having fewer than 400 bp. RV are used: (i) in gene therapy;
 CC (ii) for cloning genes; (iii) for (over) expression of proteins or RNAs
 CC and (iv) for transfection of hematopoietic stem cells (especially to
 CC impart resistance to chemotherapeutic agents). Because of their reduced

CC content of viral genes, RV are very safe (no expression of toxic or
 CC immunogenic proteins, no recombination with other viruses), have
 CC increased cloning capacity and express non-viral sequences at a high
 CC level. This sequence encodes a retroviral M4 mdr-1 protein which is
 CC described in the method of the invention.
 XX
 SQ Sequence 8630 BP; 2235 A; 2011 C; 2187 G; 2197 T; 0 other;

Alignment Scores:
 Pred. No.: 0 Length: 8630
 Score: 5857.00 Matches: 1163
 Percent Similarity: 95.01% Conservatives: 55
 Best Local Similarity: 90.72% Mismatches: 60
 Query Match: 90.48% Indels: 4
 DB: 21 Gaps: 4

US-09-672-725c-27 (1-1281) x AA224041 (1-8630)

QY 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPhetrlpLysMetGly 19
 DB 1220 ATGGATCTTGAAGGGACCGCAATGGAGGAGCAAGAAGAAGAACTTTTAACTGAAC 1279
 QY 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
 DB 1280 AATAAAGT---GAAAAGATAAGAGGAAAGAAACCAACTCTCAGTGTATTTCAATG 1336
 QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLalile 59
 DB 1337 TTTTCGGTATTCAAAATGGGCTTGCAAGTTGTATATGGTGGGGAACCTTTGGCTGCCAATC 1396
 QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
 DB 1397 ATCCATGGGCTGGACTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTT 1456
 QY 80 AlaAsnAlaGlyLysSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
 DB 1457 GCAAAATGCAGGA---AATTTAGAAGATCTGATGTCAAACATCACTAATAAGAGTGATATC 1513
 QY 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrValAlaTyrTyr 119
 DB 1514 AATGATACAGGGTCTTCATGAAT---CTGGAGCAAGACATGACAGATATGCTTATAT 1570
 QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
 DB 1571 TACAGTGAATTTGGTGGGGTCTGGTGTCTGCTTACATTCAGGCTTCATTTTGGTGC 1630
 QY 140 LeuAlaIleGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
 DB 1631 CTGGCAGCTGGAAGACAAATACAAAAATAGAAAACAGTTTTTTCATGCTATAATGCGA 1690
 QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
 DB 1691 CAGAGATAGCTGTTGTATGTCGACAGATTTGGGGAGCTTAACCCGACTTACAGAT 1750
 QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
 DB 1751 GATGCTCTAAGATTAATGAAGTATTGGTGACAAATTTGGAATGTCTTTCAGTCAATG 1810
 QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
 DB 1811 GCAACATTTTCTACCTGGGTTATAGTAGGATTTACAGTGGTGGAGCTTAACCCCTTGG 1870
 QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
 DB 1871 ATTTTGGCCATCATCTCTGTCTTGGACTGTCAGCTGCTGCTGGGCAAAAGATATCT 1930
 QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysIleGlyMetPheGlnSerIle 259
 DB 1931 TCATTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC 1990
 QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
 DB 1991 TTGGCAGCAATTAGACTGTGATTGCAATTTGGAGGACAAAGAAAGAACTTGAAGGTAC 2050

QY 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
 DB 2051 AACAAAATTTAGAAAGACTAAAGAAATTTGGATAAAGAAAGCTATTACAGCAATATT 2110
 QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
 DB 2111 TCTATAGTGTCTGCTTTCTCTCATCTATGTCATCTATGCTCTGCGCTTCTGGTATGGG 2170
 QY 320 ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer 339
 DB 2171 ACCACCTTGGTCTCTCAGGGGAATATCTATTGGACAAGTACTACTGATTTCTTTCT 2230
 QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
 DB 2231 GTATTAATTTGGGCTTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAT 2290
 QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
 DB 2291 GCAAGAGGAGCAGCTTATGAATCTTCAAGATAATTTGATATAAGCCAAAGTATTGACAGC 2350
 QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
 DB 2351 TATTCGAAGAGTGGCACAACCAAGATATAATTAAAGGAAATTTTGAATTCAGAAATGTT 2410
 QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
 DB 2411 CACTTCAGTTACCCTATCTCGAAAAAGAAATTAAGATCTTGAAGGCGCTGAACCTGAAGGTG 2470
 QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
 DB 2471 CAGAGTGGCAGAGCGGTGGCCCTGGTGGAAACAGTGGCTGTGGGAAGACAGCAACAGTC 2530
 QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
 DB 2531 CAGCTGTATGACAGGCTCTATGACCCACAGAGGGGATGTCAGTGTGTATGGACAGGAT 2590
 QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
 DB 2591 ATTAGGACATAAATTAAGGTTTCTACGGGAAATCAATTTGGTGGTGGTGGTGGTGGTGGT 2650
 QY 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
 DB 2651 GTATTTGTTGCCACCCACCATAGCTGAAACATTCGCTATGCGCGTGAAATGTCCACATG 2710
 QY 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
 DB 2711 GATGAGATTGAGAAAGCTGTCAAGGAAGCAATGCTATGCTATGCTATGCTATGCTATGCTAT 2770
 QY 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
 DB 2771 CATAAATTTGACACCTGTTGGAGAGAGAGGGGCCAGTGTGAGTGTGGTGGCAGAGAGCAG 2830
 QY 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
 DB 2831 AGGATCGCATTTGACAGTGGCTGCTGCGAACCCCAAGATCCTCTCTGCTGGATGAGGCC 2890
 QY 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
 DB 2891 ACCTCAGCCTTGGACACAGAAAGCAGTGTGTTCAAGTGGCTCTGGATGAAGCCAGA 2950
 QY 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
 DB 2951 AAAGGTGCGAGCACCATTGTGTATGCTCATGCTTTGCTACAGTTCGTAAATGCTGACGCTC 3010
 QY 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
 DB 3011 ATCGCTGGTTCGATGATGGAGTCAATTTGTGGAGAAAGAAATCATGATGATCACTCATGAAA 3070
 QY 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
 DB 3071 GAGAAAGGCAATTTACTTCAAACTTGTCAATGTCAGACAGCAGGAAATGAAAGTTCATTA 3130

QY 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
 Db 3131 GAAATCGAGCTGATGAATCCAAAGTGAAATGATGCCCTTGGAAATGCTCTCAATGAT 3190
 QY 660 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
 Db 3191 TCAAGATCCAGCTTAATAAGAAAGATCAACTCGTAGGAGTGTCCTGGATCACAGCC 3250
 QY 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
 Db 3251 CAACACAGAAAGCTAGTACCAGAGAGGCTCGGATGAAAGTATACCTCCAGTTTCCTTT 3310
 QY 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
 Db 3311 TGGAGGATTATGAAGCTAAATTAACCTGAATGGCTTATTTTGTGTGGTGATTTTGT 3370
 QY 720 AlaIleIleAsnGlyLysLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
 Db 3371 GCCATTATAATGAGGCCCTGCACACGACATTTGCAATAATATTTTCAAGAGATTATAGG 3430
 QY 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
 Db 3431 GTTTTACAGAAATGATGATCTGTAACCAACAAACGACAGAAATAGTAACTTGTGTTTCACTA 3490
 QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
 Db 3491 TTGTTTCTAGCCCTTGGAAATATTTCTTTTATTATACATTTTTCCTTCAAGGTTTCAATTT 3550
 QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
 Db 3551 GGCAAGCTGGAGAGATCCCTCAACAGCGGCTCCGATACATGTTTTCGATCCATCGCTC 3610
 QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
 Db 3611 AGACAGGATCGATGTTGGTTGATGACCTTAAACACCACTGGACATTTGACTACGAG 3670
 QY 820 LeuAlaAsnAspAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
 Db 3671 CTGCGCAATGATGCTGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGCTGTAATTACC 3730
 QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleThrGlnLeu 859
 Db 3731 CAGAAATAGCAATCTTGGGACAGAAATAATATATCTTCATCTATGTTGGCAACTA 3790
 QY 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
 Db 3791 ACACCTGTACTCTAGCAATGTGACCATCATGCAATAGCAGGAGTTGTTGAATGAAA 3850
 QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
 Db 3851 ATGTTCTCTGGACAGCACTGAAAGATAAGAAAGAACTAGAGGTGCTGGGAAGATCGCT 3910
 QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
 Db 3911 ACTGAACCAATAGAAACTCTCGAACCGTGTCTTCTTCACTCAGGACAGCAAGTTTGAA 3970
 QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
 Db 3971 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAACACACATC 4030
 QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
 Db 4031 TTTGGAAATTACATTTTCCCTTCACCCAGGCAATGATGATATTTTTCCTATGCTGGATGTTTC 4090
 QY 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
 Db 4091 CGGTTTGGAGCTACTTGTGGCACTAAACCTCATGAGCTTTGAGGATGTTCTGTAGTA 4150
 QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
 Db 4151 TTTTTCAGCTGTGCTTTTGGTGCCATGGCCGTGGGGCAAGTCAGTTTCATTTGCTCTCTGAC 4210
 QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019

Db 4211 TATGCCAAAGCCAAAATATACAGAGCCCATCATCATCATCATTTGAAAAAACCCCTTTG 4270
 QY 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
 Db 4271 ATTGACAGCTACAGCAGCGAGGCCTAATGCCGACACACATTTGAAGGAATGTACATTT 4330
 QY 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
 Db 4331 GGTGAAGTTGTTATTTCAACTATCCACCCAGCGGACATCCAGTGTGCTTCAGGACTGAGC 4390
 QY 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
 Db 4391 CTGGAGGTGAAGAGGCCAGACGCTGGCTCTGGTGGGAGCAGTGGCTGGGGAAGAGC 4450
 QY 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
 Db 4451 ACAGTGGTCCAGCTCCTGGAGCGGTTCTACGACCCCTTTGGCAGGGAAAGTGTGCTGTGAT 4510
 QY 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
 Db 4511 GGCAAAGAAATTAAGCGACTGAATGTTCACTGGCTCCGAGCACACCTGGGCATCGTGCC 4570
 QY 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
 Db 4571 CAGGAGCCCATCTGTTTGACTGCAGCATGCTGAGAACATTTGCCCTATGGAGACAAACAGC 4630
 QY 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
 Db 4631 CGGGTGGTGTCACAGGAAGAGATCTGAGGGGAGCAAGAGGAGGCCAACATACATGCCCTTC 4690
 QY 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
 Db 4691 ATCGAGTCACTGCCCTAAATAATATAGCACTAAAGTAGAGACAAAGAGAACTCAGCTCTCT 4750
 QY 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
 Db 4751 GGTGCCCAGAAACACGCAATGGCCATAGCTCGTGGCCCTGTTAGACACCCCTCATATTTTG 4810
 QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
 Db 4811 CTTTTGGATGAAGCCACGCTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAAGAGCC 4870
 QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
 Db 4871 CTGGCAAGCCAGAGAGGCGCACCTGTCATTTGATTTGCTCACCGCTGTCCACCATC 4930
 QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
 Db 4931 CAGAATGCAGACTTAATAGTGTGTTTCAGAAATGCCAGAGTCAAGGAGCATGGCAGCAT 4990
 QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
 Db 4991 CAGCAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGGTCAAGTGTCCAGGCTTGAACA 5050
 QY 1280 LysArg 1281
 Db 5051 AAGCCG 5056

Search completed: November 6, 2002, 19:26:29
 Job time : 591.645 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:39:50 : Search time 58.556 Seconds
(without alignments)
5373.600 Million cell updates/sec

Title: US-09-672-725C-27

Perfect score: 6473
Sequence: 1 MPPEGRKGSAEKNEFKMGK.....LLAQKRGYFMSVQAGAKR 1281

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-O=/cgn2_1/USPTO.spool/US09672725/runat_04112002_124339_17887/app_query.fasta_1.6435
-DB=Issued_Patents_NA -Qfmt=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFL=0
-LIST=45 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=ptc -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-USER=US09672725.ecgn_1.134_@runat_04112002_124339_17887 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -FGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTC.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5857	90.5	4646	1	US-08-181-471-2
2	5857	90.5	6505	2	US-08-793-610-5
3	5857	90.5	9318	2	US-08-793-610-6
4	5854	90.4	4669	6	5206352-3
5	5830	90.1	4669	2	US-08-752-447-1
6	5815.5	89.8	4264	2	US-08-784-649A-1
7	5815.5	89.8	4264	2	US-08-784-649A-5
8	5797	89.6	4669	2	US-08-583-276-18
9	5291.5	81.7	4233	3	US-09-120-513-1
10	5291.5	81.7	4233	4	US-09-450-105-1
11	3631	56.1	2726	1	US-08-461-823-1
12	2514	38.8	4047	2	US-08-612-734B-1

13	2482	38.3	4002	2	US-08-996-545-1	Sequence 1, Appli
14	2482	38.3	4002	2	US-08-996-545-3	Sequence 3, Appli
15	2482	38.3	4002	4	US-09-328-320-1	Sequence 1, Appli
16	2482	38.3	4002	4	US-09-328-320-3	Sequence 3, Appli
17	2445	37.8	4224	1	US-08-612-521-1	Sequence 1, Appli
18	2439.5	37.7	4800	2	US-08-612-734B-3	Sequence 3, Appli
19	2290	35.4	3924	1	US-08-395-246C-1	Sequence 1, Appli
20	2058	31.8	6143	1	US-08-612-521-3	Sequence 3, Appli
21	1881.5	29.1	3924	2	US-08-996-644-3	Sequence 3, Appli
22	1881.5	29.1	3924	3	US-09-352-552-3	Sequence 3, Appli
23	1881.5	29.1	3927	2	US-08-996-644-1	Sequence 1, Appli
24	1881.5	29.1	3927	3	US-09-352-552-1	Sequence 1, Appli
25	1854	28.6	3909	1	US-08-232-537-1	Sequence 1, Appli
26	1090	16.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
27	1090	16.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
28	782.5	12.1	2376	1	US-08-394-880B-1	Sequence 1, Appli
29	768	11.9	4781	2	US-09-001-273-1	Sequence 1, Appli
30	768	11.9	4781	4	US-08-843-459A-1	Sequence 1, Appli
31	768	11.9	4847	3	US-09-061-400-1	Sequence 1, Appli
32	750.5	11.6	5889	1	US-08-463-092B-5	Sequence 5, Appli
33	750.5	11.6	5889	2	US-08-462-109A-5	Sequence 5, Appli
34	750.5	11.6	5889	2	US-08-460-907B-5	Sequence 5, Appli
35	750.5	11.6	5889	3	US-08-463-179A-5	Sequence 5, Appli
36	750.5	11.6	5889	3	US-08-461-384B-5	Sequence 5, Appli
37	726	11.2	6140	4	US-09-439-313-536	Sequence 1, Appli
38	708	10.9	5232	4	US-08-972-927-1	Sequence 1, Appli
39	708	10.9	6082	4	US-09-439-313-535	Sequence 535, App
40	687	10.6	2061	4	US-09-061-764A-17	Sequence 17, Appl
41	684.5	10.6	2244	4	US-09-061-764A-18	Sequence 18, Appl
42	682	10.5	5011	1	US-08-463-092B-3	Sequence 3, Appli
43	682	10.5	5011	2	US-08-462-109A-3	Sequence 3, Appli
44	682	10.5	5011	2	US-08-460-907B-3	Sequence 3, Appli
45	682	10.5	5011	3	US-08-463-179A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-08-181-471-2

; Sequence 2, Application US/08181471
; Patent No. 5641508
; GENERAL INFORMATION:
; APPLICANT: Li, Lingna
; APPLICANT: Lishko, Valery K
; TITLE OF INVENTION: METHOD FOR DELIVERING BENEFICIAL
; TITLE OF INVENTION: COMPOSITIONS TO HAIR FOLLICLES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas Fitting
; STREET: 12526 High Bluff Drive, Suite 300
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92130
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/181.471
; FILING DATE: 13-JAN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/041.553
; FILING DATE: 02-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: ANT0029P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-792-3680
; TELEFAX: 619-792-8477

Db 2216 ATCGCTGGTTTCGATGATGGAGTCATTTGGAGAAAGAAATCATGATGAACATCAATGAAA 2275
Qy GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db GAGAAAGGCATTTACTTCAAACTTGTGCACATGCAGACAGCAAGAAATGAAGTTGAATTA 2335
Qy GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db GAAATCGAGCTGATGAATCCAAAGTGAATTCATGCTTCCCAATGAT 2395
Qy SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db TCAAGATCCAGTCTTAATAAGAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAAGCC 2455
Qy GluAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCCTTT 2515
Qy TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db TGGAGGATTAAGAAGCTAAATTTAACTGAATGGCTTATTTTGTGTGTGTTATTTGT 2575
Qy AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db GCCATTATAAATGGAGGCTTGCACACAGCAATTTGCAATATATATTTTCAAGATTATAGG 2635
Qy IlePheThrArgAspGluAspProGluThrLysArgGluAsnSerAsnMetPheSerVal 759
Db GTTTTACAGAAATTGATGATCTGAAACAAACAGCAAGTAACTTGTTCACATA 2695
Qy LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db TTGCTTTCTAGCCCTTGGAAATATTTCTTTTATTACATTTTCTTCAGGGTTTCACATTT 2755
Qy GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db GGCAGAGCTGGAGAGATCTCCACAGCGGCTCCGATACATGTTTTCGATCCATGCTC 2815
Qy ArgGlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db AGACAGATGTGATGTTGGTTGATGACCTTAAACACACCCTGGAGCATTTGACTACCAAG 2875
Qy LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTCTGTAATTACC 2935
Qy GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db CAGAAATATAGCAATCTTGGACAGGAATAATTATATCCTTATGTTGGCAACTA 2995
Qy ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db ACACCTTTACTCTTAGCAATTTGACCATTCATTCGAATAGCAGGAGTTGTTGAAATGAAA 3055
Qy MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db ATGTTGTCTGGAACAAGCATGAAGATAGAAAGAACTAGAAAGTGTCTGGGAAGATCGCT 3115
Qy ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db ACTGAAGCAATAGAAAACCTTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAAAGTTTAAA 3175
Qy TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db CATATGTATGCTCAGAGTTTGCAGTACCATACAGAAACTCTTTGAGGAAGACACACATC 3235
Qy PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db TTTGGAATTAACATTTCTTCCACCAGCAATGATGATTTTCTTATGCTGGATGTTTC 3295
Qy ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db CGGTTTGGAGGCTACTTGTGGCACATTAACATCATGAGCTTTGAGGATGTTCTGTAGTA 3355

Qy PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db TTTTACAGCTGTTGTTGCTTTTGGTCCATGGCGGTGGGCAAGTCAGTTTCATTTCTCTGAC 3415
Qy TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db TATGCCAAAGCCCAATAATATCAGACGCCCATCATCATGATCATTTGAAAAAACCCCTTGG 3475
Qy IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db ATTGACAGCTACAGCACGGAAGSCCTAATGCCGCAACACATTCGAAGAAATGTCACATTT 3535
Qy AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db GTTGAAGTTGTATTCAACTATATCCACCGACCGGACATCCCATGCTTTCAGGAGTGAAG 3595
Qy LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db CTGGAGGTGAAGAGGGCCAGACGCTGGCTCTGGTGGGACAGCTGGCTGTGGGAAGAGC 3655
Qy ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db ACAGTGGTCCAGCTCCTGGAGCGGTTCTAGCAGCCCTTGGCAGGGAAAGTGCTGCTTGCAT 3715
Qy GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db GGCAAGAAATAAAGCGACTGAATGTTCAAGTGGCTCCGAGCAGCACCTGGGCTATGCTGCC 3775
Qy GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db CAGGAGCCCATCTCTTGTACTGACGACATGCTGAGAACATTTGCCATGGAGAACACAGC 3835
Qy ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db CGGGTGGTGTACAGGAAGAGATCGTGAGGCGCAGCAAGAGGAGGCCAACATACATGCCTTC 3895
Qy IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db ATCGAGTCACTGCTCTTAAATAATATAGCACTAAAGTAGGAGACAAAGAACTCAGCTCTCT 3955
Qy GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db GGTGCCAGAAACACACGATTTGCCATAGCTCGTCCCTTGTGTAGACAGCCTCATATTTTG 4015
Qy LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db CTTTGGATGAAGCCACGCTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAAAGCC 4075
Qy LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db CTGGACAAAGCCAGAGAGGCGCCACCTGCAATTTGTGATTTGCTACCCGCTGTCCACCATC 4135
Qy GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db CAGAATGCAGACTTAATAGTGGTCTTTTCAGAAATGCGACAGTCAAGGAGCATGGCACCAT 4195
Qy GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db CAGCAGCTGTGTGGACAGAAAGGATCTATTTTTCAATGTTGCTGAGTCCAGGCTGGAACA 4255
Qy LysArg 1281
Db AAGCGC 4261

RESULT 2

US-08-793-610-5
; Sequence 5, Application us/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARBERS, Carol
; APPLICANT: OSTERWAG, Wolfram

```

; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
; TITLE OF INVENTION: FOR GENE TRANSFER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,610
; FILING DATE: 07-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 31 973.8
; FILING DATE: 08-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 195 03 952.1
; FILING DATE: 07-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/031175
; FILING DATE: 10-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berman, Richard J.
; REGISTRATION NUMBER: 39,105
; REFERENCE/DOCKET NUMBER: P1614-7007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6505 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
;
US-08-793-610-5
;
Alignment Scores:
Pred. No.: 0 Length: 6505
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 2 Gaps: 4
;
US-09-672-725C-27 (1-1281) x US-08-793-610-5 (1-6505)
;
Qy 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1817 ATGATCTTGAAGGGGACCGCAATGGAGGACCAAGACAGAGACTTTTAAACTGAAC 1876
;
Qy 20 LysLysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 1877 AATAAAAGT---CAAAAGATAAAGAGGAGGAAAAAGAACCAACTGTCAGTGTATTTTCAATG 1933
;
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 1934 TTTTCGCTATTCAAAATTTGGCTTGACAAGTTGTATGGTGGGGAACCTTGGCTGCCATC 1993
;
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 1994 ATCCATGGGCTGCAGCTTCCTCTCATGATCGCTGGTGGTGGGAGAAATGACAGATATCTTT 2053
;
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 2054 GCAATCTCAGGA---AATTTAGAGAGATCTGATGTCAAACATCCTAAATGAGAGTCAATC 2110
;

```

Db 3188 ATTAGGACCATAAATGTAAGGTTTCTACGGGAATAATCATTTGGTGGTGAGTCAGGAACCT 3247
QY |||||
Db 480 ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
QY |||||
Db 3248 GTATGTGTTGCCACCATGATGCTGGAACATTCGCTATGCGCGTGAATGTCACCATG 3307
QY |||||
Db 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
QY |||||
Db 3308 GATGAGATTGAGAAGCTGTCAAGGAAGCAATGCCATGACTTTATCATGAACATGCCCT 3367
QY |||||
Db 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
QY ::|||
Db 3368 CATAAATTTGACACCTGGTGGAGAGAGAGGGGCCAGTTGAGTGGTGGCGAGAAGCAG 3427
QY |||||
Db 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
QY |||||
Db 3428 AGGATGCCCATTGCACGTGCCCTGGTTGCGCAACCCCAAGATCCTCCTGCTGGATGAGGCC 3487
QY |||||
Db 560 ThrSerAlaLeuAspThrGluSerGluAlaValAlaGlnValAlaLeuAspLysAlaArg 579
QY |||||
Db 3488 ACGTGACCTGGACACAGAAAGCAAGCAGTGGTTGAGTGGCTGCTGGATGAAGCCAGA 3547
QY |||||
Db 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
QY |||||
Db 3548 AAAGGTCGACCATTTGATGATGCTCATGCTTGTCTACAGTTGCTAATGCTGACGTC 3607
QY |||||
Db 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspLeuLeuMetLys 619
QY |||||
Db 3608 ATCGCTGGTTCATGATGGAGTCATTTGGGAAAGGAATCATGATGAATCATGAAA 3667
QY |||||
Db 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
QY |||||
Db 3668 GAGAAAGCATTTACTTCAAACTGTGCAATGCAGACAGCAGGAATGAAGTTCAATTA 3727
QY |||||
Db 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
QY |||||
Db 3728 GAAATGCACTGATGAATCAAAAGTGAATGTGATGCTTGGAAATGCTTCAAAATGAT 3787
QY |||||
Db 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
QY |||||
Db 3788 TCAAGATCCAGTCTAATAAGAAAAGATCACTCTGAGAGTGTCCGTGGATCACAAGCC 3847
QY |||||
Db 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
QY |||||
Db 3848 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTCCCTTT 3907
QY |||||
Db 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
QY |||||
Db 3908 TGGAGGATATGAGCTAAATTTAACTGAATGGCCTTATTTGTTGGTGTATTTGT 3967
QY |||||
Db 720 AlaIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
QY |||||
Db 3968 GCCATTATAATGAGGCGCTGCAACACGATTTGCAATAATATTTCAAGATATAGGG 4027
QY |||||
Db 740 IlePheThrArgAspClnAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
QY ::|||
Db 4028 GTTTTACAAAGAAATGATGATCCTGAAACAAAACGACAGAATAGTAATCTGTTTCACTA 4087
QY |||||
Db 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
QY |||||
Db 4088 TTGTTTCTAGCCCTTGGAAATATTTCTTTTATTTATATTTTCTTCAGGGTTTCACATTT 4147
QY |||||
Db 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
QY |||||
Db 4148 GGCAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATGTTTCCGATCCATGCTC 4207
QY |||||
Db 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
QY |||||
Db 4208 AGACAGGATGTGAGTGTGTTGATGACCTTAAACACCACTGGAGCATTGACTTACCAGG 4267
QY |||||
Db 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
QY |||||

Db 4268 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTAATTACC 4327
QY |||||
Db 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeu 859
QY |||||
Db 4328 CAGAATATAGCAAAATCTTGGACAGGAATAATTATATCCTTCATCTATGTTGGCAACTA 4387
QY |||||
Db 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
QY |||||
Db 4398 ACATGTTTACTTCITAGCAATTTGTACCCATCATTTGCCAATAGCAGGAGTTGTTGAAATGAAA 4447
QY |||||
Db 880 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 899
QY |||||
Db 4448 ATGTTGCTCGACAGCACTGAAGATAGAAGAACTAGAAGTCTCTGGGAAGATCGCT 4507
QY |||||
Db 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
QY |||||
Db 4508 ACTGAAGCAATAGAAAACCTTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAAGTTTGA 4567
QY |||||
Db 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
QY ::|||
Db 4568 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAACACACATC 4627
QY |||||
Db 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
QY |||||
Db 4628 TTTGGAATTACATTTCCITTCACCCAGGCANTGATGATATTTTCTATGCTGGATGTTTC 4687
QY |||||
Db 960 ArgPheGlyAlaTyrLeuValAlaAlaPheMetAsnPheClnAspValLeuLeuVal 979
QY |||||
Db 4688 CGGTTTGGAGCCACTTGGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTGTA 4747
QY |||||
Db 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProasp 999
QY |||||
Db 4748 TTTTCAGCTGTTGCTTTGTCGCTGGCGGCGGCAAGTCAGTTCATTTGCTCCTGAC 4807
QY |||||
Db 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
QY |||||
Db 4808 TATGCCAAAGCCAAAATATCAGCAGCCCATCATCATGATCATTTGAAAGAAATGTCA 4867
QY |||||
Db 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
QY |||||
Db 4868 ATTGACAGCTTACAGCAGGAGCCCTTAATGCCGAACACATTTGGAAGGAATGTCA 4927
QY |||||
Db 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
QY |||||
Db 4928 GGTGAAGTTGTTTCAACTATCCACCCGACCGGACATCCAGCTCTCAGGACTGAGC 4987
QY |||||
Db 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
QY |||||
Db 4988 CTGGAGGTGAAGAAGGCCAGACGCTGCTCTGTGGCAGCAGTGGCTGTGGGAAGAGC 5047
QY |||||
Db 1080 ThrValValGluLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuLeasp 1099
QY |||||
Db 5048 ACAGTGGTCCAGCTCTCGAGCGGTTCTACGACCCCTTTGGCAGGGAAGTCTGCTTGT 5107
QY |||||
Db 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
QY |||||
Db 5108 GGCAAGAATAAAGCGACTGATGCTCAGTGGCTCCAGGACACACCTGGGCATCGTGTCC 5167
QY |||||
Db 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
QY |||||
Db 5168 CAGGAGCCCATCCTGTTTACTGTCAGCATTTGCTGAGAACAATTCCTATGGAGACAACAGC 5227
QY |||||
Db 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
QY |||||
Db 5228 CGGTTGGTGTACAGGAAGAGATCTGAGGGCAGCAAGGAGGCAACATACATGCTC 5287
QY |||||
Db 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
QY |||||
Db 5288 ATCGAGTCACTGCCCTAATAATATAGCACTAAAGTAGGACAAAGAACTCAGCTCTCT 5347
QY |||||
Db 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
QY |||||
Db 5348 GGTGCCCAAGAACACGCAATGGCATAGCTGTCGCTTGTGTAGACAGCCTCATATTTTG 5407

QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 5408 CTTTTGGATGAAGCCACGCTCAGCTGGATACAGAAAGTGAAGAGGTGTCCAAAGAGCC 5467
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 5468 CTGGACAAGCCAGAGAGCGCGACCTGCATTGTGATTGCTCACCGCGCTGTCCACCATC 5527
QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 5528 CAGAATGCAGACTTAATAGTGTGTTTCAGAAATGGCAGAGTCAAGAGCATGGCACGCAT 5587
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 5588 CAGCAGCTGCTGGCAGAGAGGCATCTATTTTTCNAATGGTCAGTGTCCAGGCTGGAACA 5647
QY 1280 LysArg 1281
Db 5648 AAGCGC 5653
RESULT 3
US-08-793-610-6
; Sequence 6, Application US/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARRERS, Carol
; TITLE OF INVENTION: OSTERTAG, Wolfram
; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
; NUMBER OF INVENTIONS: FOR GENE TRANSFER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Harmeinstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,610
; FILING DATE: 07-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 31 973.8
; FILING DATE: 08-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 195 03 952.1
; FILING DATE: 07-FEB-1995
; PRIOR APPLICATION DATA: PCT/EP95/03175
; APPLICATION NUMBER: PCT/EP95/03175
; FILING DATE: 10-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bertram, Richard J.
; REGISTRATION NUMBER: 39,105
; REFERENCE/DOCKET NUMBER: P1614-7007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9318 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
US-08-793-610-6

Alignment Scores:

Pred. No.: 0 Length: 9318
Score: 5857.00 Matches: 1163
Percent Similarity: 95.01% Conservative: 55
Best Local Similarity: 90.72% Mismatches: 60
Query Match: 90.48% Indels: 4
DB: 2 Gaps: 4
US-09-672-725c-27 (1-1281) x US-08-793-610-6 (1-9318)
QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1776 ATGGATCTTGAAGGGACCGCAATGGAGGACCAAGAAAGAAAGAACTTTTAAACTGAAC 1835
QY 20 LysLysSerLysLysLysGluLysLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 1836 AATAAAAGT---GAAAGATATAAGAAAGAAAGAAACCAACTGTCAGTGTATTTTCAATG 1892
QY 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 1893 TTTTCGCTATTCAAAATTTGGCTTGACAAGTTGTATATGTTGGTGGAACTTTGGCTGCCATC 1952
QY 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 1953 ATCCATGGGGCTGGACTTCCTCTCATGATGCTGTTGGTGGAGAAATGACAGATATCTTT 2012
QY 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 2013 GCAAATGCAGGA---AATTTAGAAGATCTGATGTCAACATCATCTAATAGAGTGATATC 2069
QY 100 AsnAsnTrpGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 2070 AATGATACAGGGTCTTCATCAAT--CTGGAGGAAGACATGACCAGGTATGCCATATTAT 2126
QY 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaIleGlnValSerPheTrpCys 139
Db 2127 TACAGTGAATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2186
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 2187 CTGGCAGCTGGGAAGCAAAATACAAAATTTAGAAAACAGTTTTCATGCTATAATGCGA 2246
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 2247 CAGGAGATAGCTGCTTTGATGTGCACGATGTTGGGGAGCTTAACACCCGCTTACAGAT 2306
QY 180 AspValSerLysIleAsnGluClyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 2307 GATGCTCTAAGATTAAATGAAGTTATTGTTGACAAAATTTGGAATGTTCTTTCAGTCAATG 2366
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 2367 GCAACATTTTTCACCTGGGTTTATAGTAGGATTTACACGCTGGTGGAACTAACCCCTTGTG 2426
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 2427 ATTTTGGCCATCAGTCTCTTCTTGGACTGTCAGTGTCTGGGCAAGAGATATACTATCT 2486
QY 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 2487 TCACTTACTGATAAAGAACTCTTACGATGCAAAAGCTGGAGAGTAGTACGTAAGAGGTC 2546
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 2547 TTGGCAGCAATTAGAACTGTGATTGCAATTTGGAGGACAAAAGAAAGAACTTGAAGGGTAC 2606
QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 2607 AACAAAAATTTAGAAAGAGCTAAAGAAATTTGGATTAAGAAAGAGCTATTACACCCATATT 2666
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 2667 TCTATAGGTGCTGCTTCTCCTGCTGATCTATGTCATCTTATGCTCTGGCCCTCTGGTATGGG 2726

Db 4887 GGTGAAGTTGTATTCAACTATCCACCACCGACATCCACAGTCTTCCAGGACTGAGC 4946
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 4947 CTGAGGCTGAAGAAGCCAGACGCTGGCTCTGGTGGCAGCAGTGGCTGGGAAGAGC 5006
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 5007 ACAGTGGTCAGCTCTCGAGGCGGTTCACGACCCCTTGGCAGGAAAGTGTCTGTGAT 5066
Qy 1100 GlyLysGlnLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 5067 GGCAGAAGAAATTAAGCGACTGAATGTTCACTGGCTCCGAGCACACTGGGCATCTGTC 5126
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 5127 CAGGAGCCATCTGTTGACTGCAGCATTTGCTGAGAACAATGGCTATGAGACAAACAGC 5186
Qy 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
Db 5187 CGGGTGGTGTACAGGAAGAGATCGTGAGGGCAGCAAGAGAGCCACATACATGCCCTTC 5246
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 5247 ATCGAGTCACTGCTTAATAATATAGCACTAAAGTAGGAGACAAAGAACTCAGCTCTCT 5306
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 5307 GGTGGCCAGAAACAACGCAATGGCATAGCTGCTGGCCCTTGTAGACAGCCTCATATTTG 5366
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 5367 CTTTGGATCAAGCCAGCTCAGCTCGGTACAGAAAGTGAAGAGCTTGCCAAAGAGCC 5426
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 5427 CTGGACAAAGCCAGAGAAGGCCACCTGCTGATTTGCTCACCGCCTGTCCACCATC 5486
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 5487 CAGANTGCACACTTAATAGTGGTGTTCAGANTGGCAGAGTCAAGGAGCATGGCAGCAT 5546
Qy 1260 GlnGlnLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 5547 CAGCAGCTGCTGGCACAGAAAGCATCTATTTTCAATGTCAGTGTCCAGGCTGGAACA 5606
Qy 1280 LysArg 1281
Db 5607 AAGCGC 5612

RESULT 4
5206352-3
; Patent No. 5206352
; APPLICANT: RonInson, Igor B.; Pastan Ira H.; Gottesman,
; Michael M.
; TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
; SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/622, 836
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 892,575
; FILING DATE: 01-AUG-1986
; APPLICATION NUMBER: 845,610
; FILING DATE: 28-MAR-1986
; SEQ ID NO:3
; LENGTH: 4669
5206352-3

Alignment Scores: 0 Length: 4669
Pred. No.: 5854.00 Matches: 1162
Score: .

Percent Similarity: 95.01% Conservative: 56
Best Local Similarity: 90.64% Mismatches: 60
Query Match: 90.44% Indels: 4
DB: Gaps: 4
US-09-672-725C-27 (1-1281) x 5206352-3 (1-4669)
Qy 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 425 ATGGATCTTGAAGGGACCGCAATGGAGGAGCAAAAGAAGAACTTTTAAACTGAAC 484
Qy 20 LysLysSerLysLysGlyLysGluLysLysProThrValSerThrPheAlaMet 39
Db 485 AATAAAAGT---GAAAAGATAGAAGAAAGAAACCAACTGTCAGTGTATTTCARTG 541
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuMetLeuValGlyThrMetAlaAlaIle 59
Db 542 TTTCGCTATTCAAAATGGCTTCACAAAGTTGTATATGGTGGGAACCTTTGGTGCCTATC 601
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 602 ATCCATGGGGCTGGACTTCCTCTCATGATGCTGCTTTGGAGAAATCACAGATATCTTT 661
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99
Db 662 GCAAATGCAGGA---AATTAGAAAGATCTGATGTCAAACTACTAATAGAGTGATATC 718
Qy 100 AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 719 AATGATACAGGGTCTTCATGAAT---CTGCAGGAAGACATGACCAGGTATGCTTAT 775
Qy 120 TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys 139
Db 776 TACAGTGAATTTGGTCTGGGGTGTCTGCTGTACATTCAGGTTCAATTTGGTGC 835
Qy 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 836 CTGGCAGCTGGGAAGCAAAATACAAAATTAGAAAACAGTTTTTTCATGCTATAATGCGA 895
Qy 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 896 CAGGAGATAGCTGCTTTGATGTGCACGATGTGGGAGCTTAACACCCGACTTACAGAT 955
Qy 180 AspValSerLysIleAsnGluClyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 956 GATGCTCTAAGATTAAACAAGTTATTTGGTGACAAAATTTGGAATGCTCTTCATGCTCA 1015
Qy 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 1016 GCACATTTTTCACCTGGGTTTATAGTAGATTTCACGCTGGTGGAGCTAACCCCTTGTG 1075
Qy 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 1076 ATTTTGGCCATCAGTCTCTTCTTGGACTGTGCTGCTGTGGGCAAAAGATATATCT 1135
Qy 240 SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 1136 TCATTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTACGTGAAGAGTC 1195
Qy 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 1196 TTGGCAGCAATTAGAATCTGATTTGATTTGGAGGACAAAAGAAAGAACTTGAAGGTAC 1255
Qy 280 AsnLysAsnLeuGluGluAlaLysGlyIleGlyLysLysAlaIleThrAlaAsnIle 299
Db 1256 AACAAAATTTAGAGAAGCTAAAAGAAATTTGGGATAAAGAAAGCTATTACACCCATAT 1315
Qy 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 1316 TCTATAGGTGCTGCTTCTCTGCTGATGATCATCTTATGCTCTGCTGCTCTGCTGATGGG 1375
Qy 320 ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer 339
Db 320 ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer 339

Db 1376 ACCACCTGGTCCTCAGGGGAATATTCATTGGACAAGTACTACTGTATTCCTTTCT 1435
Qy 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1436 GTATTAAATGGGCTTTAGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAAAAT 1495
Qy 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1496 GCAAGAGAGAGAGCTTATGAATCTTCAAGATAATTCAGATAATTAAGCAAGTATTGACAGC 1555
Qy 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1556 TAITCGAAGAGTGGGCACAAACAGATAATATTAAAGGAAATTTGGAATTCAGAAATGTT 1615
Qy 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1616 CACTTCAGTTACCACTCTCGAAGAAGTTAAGATCTTGAAGGGCTGAACCTGAAGGTG 1675
Qy 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrVal 439
Db 1676 CAGAGTGGGAGAGCGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGAGCACACAGTC 1735
Qy 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1736 CACGTGATGCGAGAGCTCTATGACCCACAGAGGGATGGTCACTGTGATGGACAGGAT 1795
Qy 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479
Db 1796 ATTAGGACCATAAATGAAGTTTCTACSGGAAATCATTTGGTGTGTGAGTCAGGAACCT 1855
Qy 480 ValLeuPheAlaThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1856 GTATTGTTGCCACCAGTAGCTGAAACATTCGCTATGCGGTGAAATGTCAACATG 1915
Qy 500 AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1916 GATGAGATTGAGAAGCTGTCAAGGAAGCCATGCCCTATGACTTTATCATGAACCTGCCT 1975
Qy 520 AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1976 CATAAATTTGACACCTGGTTGGAGAGAGAGGGGCCACGTTGAGTGGTGGGAGAGCAG 2035
Qy 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 2036 AGGATCGCCATTGACGTGCCCTGGTTCGCAACCCCAAGATCCTCCTGCTGGATGAGGCC 2095
Qy 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 2096 ACGTCAGCCTTGGACACAGAAAGCGAGAGCAGTGGTTGAGTGGCTTGGATGAAGGCCAGA 2155
Qy 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 2156 AAAGTTCGGACCAACCATTTGTAGTCTATCCTGTTGTCTACAGTTCTGTAATCCTGACGTC 2215
Qy 600 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 2216 ATCGCTGGTTCGATGATGAGTCAATGTGGAGAAAGAAATCATGATGAATCATGAA 2275
Qy 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleLeuLeu 639
Db 2276 GAGAAAGGCAATTACTTCAAACTTGTCAATGCAGACAGCAGGAATGAAGTTGAAATTA 2335
Qy 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2336 GAAATTCAGCTGATGAATCCAAAAGTGAATGATGCTTGAATGCTTCAAAATGAT 2395
Qy 660 SerGlySerSerLeuIleLysArgArgSerThrArgSerIleHisAlaProGlnGly 679
Db 2396 TCAAGATCCAGTCTAATAGAAAAGATCAACTCGTAGGAGTGTCCGTGGATCAACGCC 2455
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 699
Db 2456 CAAGACAGAAAGCTTAGTACCAGAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 2515

Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCys 719
Db 2516 TGGAGATTATGAAGCTAAATTTAACTGAATGCCCTATTATTTGTTGTTGGTGTATTTGT 2575
Qy 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2576 GCCATTATAATGGAGGCTGCAACAGCATTTGCAATAATATTTTCAAGATATTAGGG 2635
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 GTTTTTACAGAATTGATGATCCTGAAACAAACACAGAAATAGTAACCTGTTTCACTA 2695
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2696 TTGTTTCTAGCCCTTGGAAATTTATTTCTTTATATCATTTTCTCAGGGTTTCACATTT 2755
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 GCAAAAGCTGGAGAGATCCTCACCAGCGCTCCGATACATGTTTCCGATCCATGCTC 2815
Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2816 AGACAGGATGTGAGTTGGTTGATGACCTAAAAACACCACTGGAGCATTTGACTACCAGG 2875
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleIleGlySerArgLeuAlaValIleThr 839
Db 2876 CTCGGCAATGATGCTGCTCAAGTTAAGGGCTATAGTTTCCAGGCTTGTGTAATACC 2935
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2936 CAGAATATAGCAAACTCTGGGACAGGAATAATATATCTTCACTATGTTGGCACTA 2995
Qy 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2996 ACACGTGTACTCTTAGCAATTGTACCCATCATTCGAATAGCAGGAGTTGTTGAAATGAA 3055
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 3056 ATGTTGCTCGNACAGCACTGAAGATGAAGAAAGAACTAGAAAGGTGCTGGGAAGATCGCT 3115
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 3116 ACTGAAGCAATAGAAAACCTCCGAACCTGTTCTTTTGTGACTCAGGAGCAGAACTTTGAA 3175
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGTATGCTCAGAGTTTCAGGTACCATACAGAAACTCTTTGAGGAAAGCACATC 3235
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTGGAATTACATTTCTTCCACCCAGGCAATGATGATATTTTCTATGCTGGATGTTTC 3295
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 3296 CGGTTTGGAGCCTACTTGGTGGCACATAAACTCATGAGCTTGGAGGATGTTCTGTTAGTA 3355
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3356 TTTTCAGCTGTGTCTTGTGTGCTGCGGTGGGCAAGTCAGTTTCATTTGCTCCTGAC 3415
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAAGCCAAATATCAGACGCCCATCATCATGATCATGTAATGTAATGTAATGTA 3475
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTCACAGCTACAGCAGGAAGGCTTAATGCCGAACACATTTGGAAGAAATGTCACATTT 3535
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGTATTCAACTATCCACCACCGGACATCCAGTGTGCTTCCAGGACTGAGC 3595

Db	956	GA	GTCTCC	AGATTAA	TGAAGGA	TTGGT	GCAGAAA	TTTGG	GAATGTT	CTTT	CAGTCA	ATG	1015					
QY	200	Ala	ThrPhe	Phe	ThrGly	Phe	Ile	ValGly	Phe	ThrArg	Gly	Trp	Lys	Leu	ThrLeu	Val	219	
Db	1016	GCA	CAATTTT	CACTGG	TTTAT	AGTAG	AGTTT	TAC	AGTGG	TTTGG	AGACT	TA	ACC	CTT	TGTG	1075		
QY	220	Ile	Leu	Ala	Ile	SerPro	ValLeu	GlyLeu	SerAla	Ala	Ile	Trp	Ala	Lys	Ile	Leu	Ser	239
Db	1076	ATTT	TGGCCAT	CAGTCT	CTGTTCT	TGGACT	CTCAG	CTGCT	CAGCT	GTCTG	CTGGC	CAAG	ATA	CTACT	ATCT	1135		
QY	240	Ser	Phe	ThrAsp	LysGly	Leu	Leu	Ala	TyrAla	LysAla	GlyAla	ValAla	Glu	Val	259			
Db	1136	TCAT	TTACT	GATAA	AGAACT	CTCTT	AGCGT	ATGC	CAAA	AGCTGG	AGCAG	TAGT	AGCT	TA	GAAG	GC	1195	
QY	260	Leu	Ala	Ile	ArgThr	ValIle	AlaPhe	GlyGly	GlnLys	LysGlu	LeuGlu	ArgTyr	279					
Db	1196	TTGG	CAGCA	ATTAG	CACTGT	GATTC	ATTG	GAG	GACAA	AGAA	CACTTT	GA	AAAG	GTAC	1255			
QY	280	Asn	Lys	AsnLeu	GluAla	LysGly	IleGly	LysAla	IleThr	Ala	AsnIle	299						
Db	1256	AAC	AAAAAT	TTAG	AAAG	AGCTAA	AAAGAA	TTGG	TATA	AAAG	AGCTATT	TAC	AGC	CAAT	TATT	1315		
QY	300	Ser	IleGly	AlaAlaPhe	LeuLeu	IleTyrAla	SerTyrAla	LeuAlaPhe	TrpTyrGly	319								
Db	1316	TCT	ATAG	GTGCT	CTTCC	TGCTG	ATCTAT	GCTAT	CTTAT	GCTGG	CGCTT	CTG	TAT	GGG	1375			
QY	320	Thr	SerLeu	ValLeu	SerSerGlu	TyrThrIle	GlyGln	ValLeu	ThrValPhe	PheSer	339							
Db	1376	ACC	ACCTT	GTGCTCT	CAGGG	GAATAT	CTATT	GCA	AGTACT	CACTGT	ATTTCT	1435						
QY	340	Val	LeuIleGly	AlaPhe	SerIleGly	GlnAlaSerPro	SerIleGlu	AlaPhe	AlaAsn	359								
Db	1436	GTAT	TAAT	TGGGCTTT	TAGTGT	TGGCAG	GCATCT	CCA	AGCATTT	G	AGCATTT	GCA	AAAT	1495				
QY	360	Ala	argGly	AlaAlaTyr	GluIlePhe	LysIleIle	AspAsn	LysPro	SerIle	AspSer	379							
Db	1496	GCA	AGAGG	CACGCTT	ATGAA	ATCTT	CAAG	ATAAT	TGATA	ATAA	TAGC	CAAGT	TATT	GAC	AGC	1555		
QY	380	Tyr	SerLys	SerGly	HisLysPro	AspAsnIle	LysGlyAsn	LeuGluPhe	LysAsnVal	399								
Db	1556	TAT	TCGA	AGAGT	GGGC	CAAC	ACAG	ATAAT	TTA	AGG	AAATTT	TG	GAAT	TC	GA	AAAT	1615	
QY	400	His	Phe	SerTyrPro	SerArg	LysGlu	ValLysIle	LeuLysGly	LeuAsnLeu	LysVal	419							
Db	1616	CAC	TTCA	GTAT	ACC	CACTCG	AAAG	AGTTA	AGTCTT	TGA	AGG	CCCTG	AAC	CTG	AGG	CTG	1675	
QY	420	Gln	SerGly	GlnThrVal	AlaLeu	ValGly	AsnSerGly	CysGlyLys	SerThrVal	439								
Db	1676	CAG	AGTGGG	CAGACG	GTGGCT	TGGT	TGA	ACAGTGG	CTGTGG	GAAG	AGAC	CA	CAAC	AGTC	1735			
QY	440	Gln	LeuMet	GlnArg	LeuTyrAspPro	ThrAspGlyMet	ValCysIle	AspGlyGln	Asp	459								
Db	1736	CAG	CTGATG	CAGAGG	CTCTAT	GC	CCCA	CAGAGG	GGTGT	CA	GTGT	GTAT	GATG	AG	CA	AGG	1795	
QY	460	Ile	argThrIle	AsnValArg	HisLeu	ArgGlu	IleThrGly	ValValSer	GlnGluPro	479								
Db	1796	ATT	AGG	ACCATA	ATAAT	TAA	AGTTCTT	AC	GGG	AAATCAT	TTG	GTGT	GTGT	AGT	CA	AG	1855	
QY	480	Val	LeuPhe	AlaThrThrIle	AlaGlu	AsnIle	argTyrGly	ArgGlu	AsnValThrMet	499								
Db	1856	GTAT	TTGTTGCC	ACCACG	TAGCT	AAAA	CAATTC	GCCTAT	GGCGT	GAAT	GTG	GAAT	GTG	CA	CA	ATG	1915	
QY	500	Asp	GluIleGlu	LysAlaVal	LysGluAla	AsnAlaTyr	AspPheIleMet	LysLeuPro	519									
Db	1916	GAT	GACAT	TTGAG	AACTGCT	CA	AGG	AGCCAA	TGCCCTAT	GACT	TTAT	CA	TAT	GA	AACT	1975		
QY	520	Asn	LysPhe	AspThrLeu	ValGlyGlu	ArgGlyAla	GlnLeuSerGly	GlyGlnLysGln	539									
Db	1976	CATA	AAATTTG	AC	ACCTTGG	TGGAG	AGAGG	GGCCAG	TGTG	AGT	GGTGGG	CA						

Db	2036	AGGATCGCCATTGTCACGCTGCCTGGTTCCGAAACCCCAAGATCCCTCCTCGCTGGATGAGGCC	2095
QY	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579
Db	2096	ACGTCACGCTTGGACACAGAAAGCGAAGCAGTGGTTCCAGTGGCTCTGGATAGGCCAGA	2155
QY	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599
Db	2156	AAAGTCGGACCACCAATTGTGATAGCTCATCGTTTCTACAGTCTGTAATGCTGACGTC	2215
QY	600	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	619
Db	2216	ATCGCTGGTTTCGATGATGGAGTCATTGTCGAGAAAGAAATCATGATGAATCATGAAA	2275
QY	620	GluLysGlyIleTyrPheLysLeuValThrMetClnThrArgGlyAsnGluIleGluLeu	639
Db	2276	GAGNAAGGCATTTACTTCAAACTTGTCCAAATGCAGACACAGCAAAATGAAGTTGAATTA	2335
QY	640	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	659
Db	2336	GAATAATGCAGCTGATGAATCCAAAGTGAAATTGATGCCTTGGAAATGCTTCAAAATGAT	2395
QY	560	SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly	679
Db	2396	TCAAGATCCAGTCTTAATAAGAAAGATCAACTCGTAGGAGTCCGCTGGATCACAAGCC	2455
QY	680	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	699
Db	2456	CAACACAGAAGCTTAGTACCAAGAGGCCTCGATGAAGATATACCTCCAGTTTCCCTT	2515
QY	700	TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	719
Db	2516	TGCAGGATTATGAAGCTAAATTTAACTGAATGCCTTATTTCTGTTGTGTATTATTGT	2575
QY	720	AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	739
Db	2576	GCCATTATAATGGAGGCCTGCACCAAGCATTTTGCATAATAATTTTCAAAAGATTATAGG	2635
QY	740	IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal	759
Db	2636	GTITTTTACAGAATGATGATCTCGAAACAAAACGACAGAAATAGTAAGTTGTTTTCACAT	2695
QY	760	LeuPheLeuValLeuGlyIleIleSerPheIleThrPheLeuGlnGlyPheThrPhe	779
Db	2696	TTGTTTCTAGCCCTTGGAAATTATTCTTTTATTACATTTTCTCTCAGGGTTTCACATTT	2755
QY	780	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	799
Db	2756	GGCAAAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATGGTTTTCGGATCCATGCTC	2815
QY	800	ArgGlnAspValSerTyrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg	819
Db	2816	AGACAGGATGTGATGTGGTTTCATGACCCCTAAACACCCACTGGACATTGTACTACGAG	2875
QY	820	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	839
Db	2876	CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTCTGCTGAATTACC	2935
QY	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	859
Db	2936	CAGAATATAGCAAACTCTGGGACAGGAATTAATATATCCTTCATCTATGTTGGCAACTA	2995
QY	860	ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	879
Db	2996	ACACTGTTACTCTTAGCAATTTGATCCCATCTGCAATAGCAGGAGTTGTTGAAATGAAA	3055
QY	880	MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla	899
Db	3056	ATGTTTGTGGACACAGCACTGAAACATAAGAAAGAACTAGAGGTGCTGGGNAAGATCGCT	3115
QY	900	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu	919
Db	3116	ACTGAAGCAATAGAAACCTCCGAAACCGTTGTTCTTTGACTCAGGACAGCAAGTTTCAA	3175

Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGATGCTCAGAGTTTGAGGTACCATACAGAAATCTTTTGGAGAAACACACATC 3235
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTGGAAATACATTTTCCCTACCCAGGCAATGATGATATTTTCTATGCTGGATGTTTC 3295
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 3296 CGGTTTGGAGCCTACTTGGTGCCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTA 3355
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerPheAlaProasp 999
Db 3356 TTTTTCAGCTGTTGCTTTGGTGCCATGCGCGTGGGCAAGTCAGTTCATTTGCTCTGTCGAC 3415
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAGCCAAATATACAGAGCCACATCATCATGATCATTTGAAAAAACCCCTTTG 3475
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlnValThrPhe 1039
Db 3476 ATTGACAGCTACAGCAGGAAGCCCTAAATGCCGAACACATTTGGAAGAAATGTCACATTT 3535
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGATTAACATATCCACCACCCAGCCAGCATCCAGTGGTTCAGGAGCTGAGC 3595
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValIglySerSerGlyCysGlyLysSer 1079
Db 3596 CTGGAGGTGAAGAAGGCCACAGCGTGGCTCTGTTGGGACAGAGTGGCTGTGGGAAGAGC 3655
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3656 ACAGTGGTCCAGCTCTCGGAGCGGTCTACGACCCCTTGGCAGGAAAGTGCCTGCTGAT 3715
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 GGCAAGAAATAAAGCGACTGAATGTTTCAGTGCTCCGAGCACACCTGGGCACTGCTGC 3775
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 CAGGAGCCCATCTCTTTGTACTGCAGCATTTGCTGAGAACATTCCTATGGAGACAACAGC 3835
Qy 1140 ArgValValSerHisGluGluIleValGlnAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3836 CGGTGTGTGCACGGAAGAGATCGTGAGGCGACAAAGAGGCCCAACATACATGCGCTTC 3895
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCCAGTCACTCGCTTAATAATATAGCACATAAGTAGGAGACAAAGAACTCAGCTCTCT 3955
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GGTGGCAGAAACAACCATTTGCCATAGCTGCTGCGCTTTGTAGACAGCCCTCATATTTTG 4015
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTGGATGAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGGTTGTCCAAGAGCC 4075
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGGACAAAGCCACAGAAAGCCGACCTGCAATGCTGATGCTCACCCGCTGTCCACCATC 4135
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGAAATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGATCAAGGAGCATGCACGCAT 4195
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGCTGGCACAGAAAGGCATCTATTTTTCATGTCAGTGTCCAGGCTGGAACA 4255

Qy 1280 LysArg 1281
Db 4256 AAGCGC 4261
RESULT 6
US-08-784-649A-1
: Sequence 1, Application US/08784649A
: Patent No. 5830697
: GENERAL INFORMATION:
: APPLICANT: Sikic, Branimir I
: APPLICANT: Chen, Gang
: TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
: TITLE OF INVENTION: CYCLOSPORIN MODULATION
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Fish & Richardson
: STREET: 2200 Sand Hill Road
: CITY: Menlo Park
: STATE: CA
: COUNTRY: USA
: ZIP: 94025
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/784,649A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Sherwood, Pamela J
: REGISTRATION NUMBER: Reg.No. 5830697 36,677
: REFERENCE/DOCKET NUMBER: 06037/007001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-322-5070
: TELEFAX: 415-854-0875
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4264 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-08-784-649A-1
Alignment Scores:
Pred. No.: 0 Length: 4264
Score: 5815.50 Matches: 1162
Percent Similarity: 94.93% Conservative: 55
Best Local Similarity: 90.64% Mismatches: 60
Query Match: 89.84% Indels: 6
DB: 2 Gaps: 5
US-09-672-725C-27 (1-1281) x US-08-784-649A-1 (1-4264)
Qy 1 MetAspProGluGlyClyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 143 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAAAGAACTTTTAAACTGAAC 202
Qy 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet 39
Db 203 AATAAAGT--GAAAAGATGAAGGAAAGAAACCACTGTCAGTGTATTTCAATG 259
Qy 40 PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIle 59
Db 260 TTTTCGATTCAAAATG-CTTGACAAGTTGTATGTTGGTGGAACTTTGGCTCCCATC 318
Qy 60 IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe 79
Db 319 ATCCATGGGCGGAGCTTCCTCTCATGATGCTGGTGTGGAGAAATGCAGATATCTTT 378
Qy 80 AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr 99

QY	820	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	839	QY	1180	GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu	1199
Db	2590	CTCGCAATGATGCTCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGCTGTAATACC	2649	Db	3670	GTGGCCAGAAACACCACTGTCATAGCTGCTGCTTGTAGACAGCTCATATTTTG	3729
QY	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeu	859	QY	1200	LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla	1219
Db	2650	CAGATATAGCAAACTTTGGGACAGGAATAATATATCTTCATCTATGTTGGCAACTA	2709	Db	3730	CTTTTGGATGAAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAAGGCC	3789
QY	860	ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	879	QY	1220	LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle	1239
Db	2710	ACACTGTTACTCTTACCAATTGTACCCATCATGCAATAGCAGAGTGTGTAAGTAA	2769	Db	3790	CTGGCAAGCCAGAGAGAGCCGACCTGTCATGTTGTCACCCCTGTCCACCATC	3849
QY	880	MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluAlaGlyLysIleAla	899	QY	1240	GlnAsnAlaAspLeuIleValIlePheGlnAsnGlyLysValLysGluHisGlyThrHis	1259
Db	2770	ATGTTGCTGGACAGCACTGAAAGATAAGAAAGAACTAGAAAGTGTGGGAATCGCT	2829	Db	3850	CAGATGTCAGACTTAATAGTGTGTTTCAGAAATGGCAGACTCAAGGAGCATGGCACCAT	3909
QY	900	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu	919	QY	1260	GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla	1279
Db	2830	ACTGAAGCAATAGAAACTTCCGAACCGTGTCTTTGACTCAGGAGCAGAAATTTGAA	2889	Db	3910	CAGCAGCTGCTGGCAGCAAGAAAGGCATCTATTTTCAATGCTCAGTGTCCAGGCTGGAACA	3969
QY	920	TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle	939	QY	1280	LysArg 1281	
Db	2890	CATATGATGCTCAGAGTTGAGGTACCATACAGAACTCTTTGAGGAAAGCACACATC	2949	Db	3970	AAAGCC 3975	
QY	940	PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe	959	RESULT 7			
Db	2950	TTTGGAAATACATTTCCCTTCCACCAAGCAATGATGATTTTCTATGCTGGATGTTTC	3009	US-08-784-649A-5			
QY	960	ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal	979	; Sequence 5, Application US/08784649A			
Db	3010	CGGTTTGGAGCCCTACTTGGTGGCACAATAACTCATGAGCTTTGAGGATGTTCTGTAGTA	3069	; Patent No. 5830697			
QY	980	PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp	999	; GENERAL INFORMATION:			
Db	3070	TTTTACGCTGTTCTTGTGTCATGGCGTGGGCAAGTCAGTTCATTTCTCTCTGAC	3129	; APPLICANT: Sikic, Branimir I			
QY	1000	TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu	1019	; APPLICANT: Chen, Gang			
Db	3130	TATGCCAAAGCAAAATATCAGCAGGCCACATCATCATGATCATTTGAAAGAAATGTCACATTT	3189	; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO			
QY	1020	IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe	1039	; TITLE OF INVENTION: CYCLOSPORIN MODULATION			
Db	3190	ATTACAGCTACACACAGGAGGCTTAATGCCGACACATTTGGAAGAAATGTCACATTT	3249	; NUMBER OF SEQUENCES: 5			
QY	1040	AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer	1059	; CORRESPONDENCE ADDRESS:			
Db	3250	GGTGAAGTTGTATTCAACTATCCACCACCGGACATCCAGTGTCTCAGGAGCTGAC	3309	ADDRESSER: Fish & Richardson			
QY	1060	LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer	1079	STREET: 2200 Sand Hill Road			
Db	3310	CTGGAGGTGAAGAGGGCCAGCCTGCTGCTGGGCGCAGCTGCTGGTGGGCAAGTGGCTGGGAAGAGC	3369	CITY: Menlo Park			
QY	1080	ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp	1099	STATE: CA			
Db	3370	ACAGTGGTCCAGCTCTCTGGAGGGTTCTTACGACCCCTTGGCAGGAAAGTGGCTGTGAT	3429	COUNTRY: USA			
QY	1100	GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer	1119	ZIP: 94025			
Db	3430	GGCAAGAAATAAGCGACTGAATGTTTCAGTGGCTCCGACACACACCTGGGCATCGTGTCC	3489	COMPUTER READABLE FORM:			
QY	1120	GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer	1139	MEDIUM TYPE: Floppy disk			
Db	3490	CAGGAGCCCATCTCTTTGACTGTCAGCATTTGCTGAGAACATGCTGATGAGACACAGC	3549	COMPUTER: IBM PC compatible			
QY	1140	ArgValValSerHisGluGluIleValGlnAlaLysGluAlaAsnIleHisHisPhe	1159	OPERATING SYSTEM: PC-DOS/MS-DOS			
Db	3550	CGGGTGGTGTACAGAGAGAGATTGTGAGGAGCAGCAAGAGGCGCCACATACATGCTTC	3609	SOFTWARE: PatentIn Release #1.0, Version #1.25			
QY	1160	IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer	1179	CURRENT APPLICATION DATA:			
Db	3610	ATCGAGTCTACTGCTTAATAATATAGCACTAAAGTAGGAGACAAAGGAAGTCTCTCT	3669	APPLICATION NUMBER: US/08/784,649A			
				FILING DATE:			
				CLASSIFICATION: 435			
				ATTORNEY/AGENT INFORMATION:			
				NAME: Sherwood, Pamela J			
				REGISTRATION NUMBER: Reg.No. 5830697 36,677			
				REFERENCE/DOCKET NUMBER: 06037/007001			
				TELECOMMUNICATION INFORMATION:			
				TELEPHONE: 415-322-5070			
				TELEFAX: 415-854-0875			
				INFORMATION FOR SEQ ID NO: 5:			
				SEQUENCE CHARACTERISTICS:			
				LENGTH: 4264 base pairs			
				TYPE: nucleic acid			
				STRANDEDNESS: single			
				TOPOLOGY: linear			
				MOLECULE TYPE: cDNA			
				US-08-784-649A-5			
				Alignment Scores:			
				Pred. No.:	0	Length:	4264
				Score:	5815.50	Matches:	1162
				Percent Similarity:	94.93%	Conservative:	55
				Best Local Similarity:	90.64%	Mismatches:	60
				Query Match:	89.84%	Indels:	6
					2	Gaps:	5

US-09 -672-725C-27 (1-1281) x US-08-784 -649A-5 (1-4264)

Qy	1	MetAspProGluGlyArgLysGlySerAla-- --GluLysAsnPheTrpLysMetGly	19
Db	143	ATGGATCTGAAGGGACCGCAATGGAGGAGCAAGAAGAACCTTTTAAACTGAAC	202
Qy	20	LysLysSerLysLysGluLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMet	39
Db	203	AATAAAGT-- --GAAAAAGATAGAAGGAAAAGAACCAACCACCTGTCTAGTGATTTCATG	259
Qy	40	PheArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle	59
Db	260	TTTCCTATTCAAATG-CTTGACAAGTTGTATATGGTGGGAACCTTTGGCCCATC	318
Qy	60	IleHisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPhe	79
Db	319	ATCCATGGGGCTGGACTTCTCTCATGATGCTGGTGTGGAGAAATGACACATATCTTT	378
Qy	80	AlaAsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThr	99
Db	379	GCAATGCAGGA-- --AATTTAGAAGATCTGATGTCACACATCATCAATAGAAAGTAGATATC	435
Qy	100	AsnAsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr	119
Db	436	AATGATACAGGGTCTTCATGNAAT-- --CTGAGGAGAACATGACACAGTATGCCCTATTAT	492
Qy	120	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys	139
Db	493	TACAGTGAATTTGGTCTGGGGTCTGGTGTCTGCTTACATTCAGGTTTCATTTTTGGTGC	552
Qy	140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159
Db	553	CTGCAGCTGGAGACAAATACAAAATTAAGAAACAGTTTTTTCATGCTATATATGCGA	612
Qy	160	GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179
Db	613	CAGGAGATAGGCTGGTTGATGTCACCATGTTGGGAGCTTAACACCGACTTACAGAT	672
Qy	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheIleSerIle	199
Db	673	GATGCTCCAAGATTAATGAAGGAATTTGGTGACAAAATTTGAATG-- --TTCACGTCATAT	729
Qy	200	AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219
Db	730	GCAACATTTTTCATCTGGSTTTATAGTAGGATTTACACGTGGTTGGAAGCTAACCCCTTGTG	789
Qy	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239
Db	790	ATTTTGGCCATCAGCTCTGTTCTTGGACTGTCTAGCTGCTGTCTGGGCAAGATACTATCT	849
Qy	240	SerPheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal	259
Db	850	TCATTTACTGANAAGAACCTTAGCCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTC	909
Qy	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr	279
Db	910	TTGGCAGCAATTAGACCTGTGATTGCAATTTGGAGCACAAAGAAAGACITGAAGGTAC	969
Qy	280	AsnLysAsnLeuGluGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299
Db	970	ACAAAAAATTTAGAAGAAAGCTAAAGAAATTTGGATAAAGAAAGCTATTACAGCCAAATTT	1029
Qy	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319
Db	1030	TCATATAGTCTGCTTCTTCCTGCTGATCATGCTATGCTCTGGCCCTCTGGTATGGG	1089
Qy	320	ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer	339
Db	1090	ACCACCTTGGTCTCTCAGGGGAATATCTATTGGACAAGTACTCACTGTATTTCTTCT	1149
Qy	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359

|||||
Db 2216 ATCGCTGGTTTCGATGATGGAGTCAATGTGGAGAAAGGAATCATGATGAACATCATGA 2275
Qy 620 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 2276 GAGAAAGGCATTACTTCAAACTTGTCAATGCACAGCAGGAGAAATGAAGTTGAATTA 2335
Qy 640 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 659
Db 2336 GAAATATGCAGCTGATGAATCCAAAAGTGAATGTAGTGGTGGAAATGCTTCAAAATGAT 2395
Qy 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleIleHisAlaProGlnGly 679
Db 2396 TCAAGTACAGCTTAATAGAAAAGATCACTCGTAGAGTGTCCGTGGATCAACAACC 2455
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLysLeuAsnGluAsnValProProValSerPhe 699
Db 2456 CAAGACAGAAAGCTTAGTACCAAGAGGCTTGGATGAAGTATACCTCCAGATTTTCCTTT 2515
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2516 TGGAGGATTATGAAGCTAAATTAACCTGAATGGCCTTATTTTGTGTGGTGTATTTGT 2575
Qy 720 AlaIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2576 GCCATTATAATGGAGGCTGCAACCCAGCATTTGCAATTAATATTTCAAAGATTATAGGG 2635
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2636 GTTTTTCACAGAATTGATGATCCTGAAACAAAACGACAGAACTAACTGTGTTTTCACATA 2695
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2696 TTGTTTCTAGCCCTTGGAAATTATTTCTTTTATTTACATTTTCTTCAGGGTTTCACATTT 2755
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2756 GGCAAAAGCTGGAGAGATCCTCACCAGCGCTCCGATACATGGTTTTCGATCCATGCTC 2815
Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2816 AGACAGATGTAGTTGGTTGTATGACCCCTAAACACCCACCTGGAGCATGTACTACCAG 2875
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2876 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCCAGGCTGCTGTAATTACC 2935
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2936 CAGAAATATAGCAATTTCTGGGACAGGAATAATATATCTTCATCTATGGTTGGCAACTA 2995
Qy 860 ThrLeuLeuLeuAlaIleValProIleAlaIleAlaIleValValGluMetLys 879
Db 2996 ACACGTGTACTCTTAGCAATGTACCACTCAATTTGCAATAGCAGAGTGTGTAATGAAA 3055
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluIleGlyLysIleAla 899
Db 3056 ATGTGTCTGGACACGACCTGAAAGATAAGAAAGAACTAGAAAGTGTCTGGAAAGATCGCT 3115
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 3116 ACTGAAGCAATAGAAAATCTCCGAAACCGTTGTTCTTTGACTCAGGAGCAAGATTGAA 3175
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 3176 CATATGATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAGACACACATC 3235
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 3236 TTTTGGAAATTACATTTCTCTCCACCGCAATGATGATTTTCTTATGTTGATGATGTTTC 3295
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
|||||

Db 3296 CGGTTTGGAGCCCTACTGTGGGCACATAAACTAATAGAGCTTTCAGGATGTTCTGTAGTA 3355
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3356 TTTTCAGCTGTTGTCITTTGTGGCATGGCGTGGCAAGTGAAGTTCATTTGCTCCTGAC 3415
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3416 TATGCCAAAGCCAAAATATCAGCAGCCACATCATCATCATGATTTGAAAACCCCTTTG 3475
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3476 ATTGACAGCTACACGCGAAGGCTTAATGCCGAACACATTTGGAAGAAATGTTCACATTT 3535
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3536 GGTGAAGTTGATTCAACTATCCACCACCGACATCCAGTCTCTTCAGGAGCTGAGC 3595
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3596 CTGGAGGTGAAGAGGCGCAGACGCTGGCTGTGGTGGCAGCAGTGGCTGTGGGAAGAGC 3655
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3656 ACAGTGGTCCAGCTTCCTGGAGCGGTTCTACGACCCCTTGGCAGGAAAGTGTCTTGAT 3715
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3716 GGCAAAAGAAATAAGCGACTGAATGTTCACTGGCTCCGAGCAGCACACCTGGCATGCTGC 3775
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3776 CAGGAGCCCATCTCTGTTGACTGCGAGCATTTGCTGAGAACATTTGCTATGGACACACAGC 3835
Qy 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1159
Db 3836 CGGTGGTGTGCACAGGAAGAGATCGTGAGGSCACAAAGGAGGCCAACATACATGCCTTC 3895
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3896 ATCGAGTCACCTGCCTTAATAATATAGCATAAGTAGGAGACAAAGGAACCTCATCTCT 3955
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3956 GTGGCCAGAAACACGCAATTTGCCATAGCTCGTGCCTTGTAGACAGCCTCATATTTTG 4015
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 4016 CTTTTGGATGAAGCCACCTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAAGAGCC 4075
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 4076 CTGGCAAAACCCAGAGAGGCGCACCTGCAATTTGATTGCTCACCCTGTCCACCATC 4135
Qy 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 4136 CAGATGACAGACTTAATAGTGTGTTTCAGAAATGSCAGAGTCAAGCAGCATGGCAGCAT 4195
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 4196 CAGCAGCTGCTGGCAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGCCTGGAAACA 4255
Qy 1280 LysArg 1281
Db 4256 AAGCGC 4261
RESULT 9
US-09-120-513-1
; Sequence 1, Application US/09120513
; Patent No. 6025160
; GENERAL INFORMATION:
; APPLICANT: Brun, Kimberly
; APPLICANT: Chenery, Richard

APPLICANT: Ellens, Harma
 APPLICANT: Field, John
 APPLICANT: Yue, Lin
 TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE
 TITLE OF INVENTION: SEQUENCES ENCODING RAT MDR1B2 AND
 TITLE OF INVENTION: SCREENING METHODS THEREOF
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: SmithKline Beecham Corporation
 STREET: 709 Swedeland Road
 CITY: King of Prussia
 STATE: PA
 COUNTRY:
 ZIP: 19406

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/120,513
 FILING DATE: 22-JUL-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: King, William T

REGISTRATION NUMBER: 30,954

REFERENCE/DOCKET NUMBER: GP50008

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5015

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4233 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

US-09-120-513-1

Alignment Scores:

Pred. No.:	0	Length:	4233
Score:	5291.50	Matches:	1030
Percent Similarity:	90.48%	Conservative:	130
Best Local Similarity:	80.34%	Mismatches:	113
Query Match:	81.75%	Indels:	9
DB:	3	Gaps:	5

US-09-672-725C-27 (1-1281) x US-09-120-513-1 (1-4233)

QY	1	MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys	20
Db	26	ATGGAGTTTGAAGAGGGCTTAACGGACAGCAGACAGAACTTCTCAAGATGGGCAAA	85
QY	21	LysSerLysLysGluLysGlyLysProThrValSerThrPheAlaMetPhe	40
Db	86	AGAGTAAAG-----GAGAAGGAGAGAAACCTGCTTGGCAVATTCGGATGTT	139
QY	41	ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle	60
Db	140	CGCATGCGAGATGGCTTGACAGCTGCGATGGCTCGGGAACCTCGCTGCTATCATC	199
QY	61	HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	80
Db	200	CACGGAAACCTCTTCCCTCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	259
QY	81	AsnAlaGlyLysSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	100
Db	260	CAAGCA-----GAGACCCGATTCCTCCGAGCGTTACTAATCAAGTGAATCAAC	310
QY	101	AsnThrGlnHisPheIle---AsnHisGluGluGluMetThrThrTyrAlaTyrTyr	119
Db	311	ACTACACAGACCGTCAGGACGAGCAGTCTGGAGGAGACATGGCCATGTACGCTACTAT	370

QY	120	TyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCys	139
Db	371	TACACGGGCATTTGGTGGCGGTGCTCTCATCTGTTGCCCTACATCAGGTTTCTGCTGC	430
QY	140	LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg	159
Db	431	CTGGCAGCTGGGAGACAAATACACAAAGATTAGGCAGAAAGTTTTCATGCCATCATGAAT	490
QY	160	GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp	179
Db	491	CAGCAGATAGCTGTTTGGCTGAGTGAATGAGCTGGGAGCTCAACACCGCTCACAGAT	550
QY	180	AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPheGlnSerIle	199
Db	551	GACGCTCCAAATTAATGACGGAATTTGGTGCACAACTTGAATGTTCTTTCAGTCCATA	610
QY	200	AlaThrPheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal	219
Db	611	ACGACATTTTCAGCCGGTTTTTAATAGGATTTAAGTGGTGGAGGTAACCTCTGTA	670
QY	220	IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer	239
Db	671	ATTTTGGCGCTCAGCCCTCTTATTTGGTGTCTCATCTGCCATGTGGCAAGGACTGACT	730
QY	240	SerPheThrAspLysGluLeuAlaTyrAlaLysAlaGlyAlaValAlaGluVal	259
Db	731	TCATTTTACTAATAAGGAACCTCCAGGCTTATCGAAAGCTGGAGCAGTTGCCGAAGATC	790
QY	260	LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuArgTyr	279
Db	791	TTAGCAGCAATCAGAACTGTGATTTGGTGGTGGAGGCAAAAGAAAGAACTTGAAGGTC	850
QY	280	AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle	299
Db	851	AATAAAATTTAGAGAAGCTAAAGAGTTGGCATAAGAAAGCCATCAGGCCAACATT	910
QY	300	SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly	319
Db	911	TCCATAGGTATTCCTACCTGTTGGTGTCTATGCTTATGCTGCTGCTGCTGCTGCTGCT	970
QY	320	ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer	339
Db	971	ACCTCCTTGGTCTCTCAATGATTCATTTGACAGAGCTTACCGCTTCTCTCTCT	1030
QY	340	ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn	359
Db	1031	ATTTTATTTGGGACTTTCAGTATTTGACATTTAGCCCCAACATAGAACCTTTTGCAAT	1090
QY	360	AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer	379
Db	1091	GCAGAGGGGCGAGCCTATGAATCTTCAAGATAATTTGATAATGAGCCCAAGCATTTGACAGC	1150
QY	380	TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal	399
Db	1151	TTCTCAACCAAGGACGACAAACACACAGATATATGGGAAATTTGGAATTTAAAAATGTT	1210
QY	400	HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal	419
Db	1211	TACTTCACTCACTCCATCAGAGTGAAGTAAAGATCTTGAAGGCGCTCAACCTCAAGGTG	1270
QY	420	GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal	439
Db	1271	AGAGCGGCGACAGCGTACGCTTGGTGGCAACAGTGGCTGGTGGGAAAGCAACCACTGTC	1330
QY	440	GlnMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp	459
Db	1331	CAGCTGCTGCAGAGCTCTACGACCCCATAGAGGCGAGGTTCAGTATCGAGGACAGGAC	1390
QY	460	IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro	479
Db	1391	ATCAGGACCATCAATGTGAGGTATCTCGGGAAATCATTTGGGTGGTGTGAGTACGAGAACCC	1450

Qy	480	ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet	499
Db	1451	GTGCTGTTGGCCACACGATTCGCCAAAACATTCGCTATGGCGAGAAAAACGTCACCATG	1510
Qy	500	AspGluIleGluLysAlaValLysGluAlaAsnAlaThrAspPheIleMetLysLeuPro	519
Db	1511	GATGAGATAGAAAGACTCTCAAGGAAGCCAAATGCCATGACTTCATCATGAACATGGCCC	1570
Qy	520	AsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln	539
Db	1571	CACAAATTTAACCCCTGTTGGTGAGAGGGCGCAGCTGAGTGGGGACAGAAACAG	1630
Qy	540	ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla	559
Db	1631	AGGATGCCCATTCGCCGGGCCCTGGTCCGCACACCCCAAGATCCCTTTGTTGGATGAGGC	1690
Qy	560	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	579
Db	1691	ACGTGACGCTTGGACACAGAAACGAGCGGTTCAGGCGGCTCTGGATAAGGCTAGA	1750
Qy	580	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	599
Db	1751	GAAGCGGACCAACCATTTGTATGATCTACCCGTTGTCTACAGATGCCAAATGCTGAGTC	1810
Qy	600	IleAlaGlyPheAspGlyValIleValGluLysGlyLysAsnHisAspGluLeuMetLys	619
Db	1811	ATTGCTGGTTTGATGGTGTGTCATTGTGGAGCAAGGAATCATGAAGAGCTCATGAAA	1870
Qy	620	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu	639
Db	1871	GAGAAGGCAATTTACTTCAAACTTGTGATGCACACAGACTAGAGGAAATGAATTTGAACCA	1930
Qy	640	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	659
Db	1931	GGAATAATGCTTATGAATCCCAAGATGACACTGGTGCCCTGAGTGTGACTTCAGAAAA	1990
Qy	660	SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly	679
Db	1991	TCAAATCTCCTTTAATA--AGGAGATCAATTCGCAGAAGTATCCACAGAAGACAAGAC	2047
Qy	680	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	699
Db	2048	CAGGAGAGAAGACTTAGTTCGAAGAGGATGGATGAAGATGTGCCATATGTTTCCCTTT	2107
Qy	700	TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	719
Db	2108	TGGCAGATCCTTAAGCTAAATATTAGTGAATGGCCCTATTATTGTTGGTGGTACTTTGT	2167
Qy	720	AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	739
Db	2168	GCTGTTATTAATATGGTGTCATACACCACTGTTGGCATAGTGTTCGAAGATTTGATGGG	2227
Qy	740	IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal	759
Db	2228	GTITTTTTCAAGACAGCAGCACCATGAACCAACACGGAATGTGAACCTGTGTTCCCTT	2287
Qy	760	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	779
Db	2288	CTCTTCTGGTCATGGGAATGATTTCTTTTGTACGTACTCTTTTCAAGGCTTCACATTT	2347
Qy	780	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	799
Db	2348	GGCAAGCTGGAGAGATCCTCACCACGCACTCCGATACATGGCTTTCAAAATCCATGCTG	2407
Qy	800	ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg	819
Db	2408	CGACAGGATATAAGCTGGTTTGTATGACCATAAAAACACCACTGGCTCGCTACTACCAGG	2467
Qy	820	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	839
Db	2468	CTCGCTAGTACGCTCTTAATGTAAAGGGGGCTATGGGCTCCAGGCTTCTGTGATTACC	2527
Qy	840	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	859

```

Db 3608 CTTCTGGATGAAGCGACATCATGCTGTGATACGAGAGTGAAGAAGTCTGCCAGGAAGCG 3667
QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3668 CTGGACAAACCCAGGAGCCGACCTGTGATTCGCGCACCAGCTGTCCACCATC 3727
QY 1240 GluAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3728 CAGAAGCAGACTGTGCTGGTGATTCAGAACGCCAGGTCAAGGAGCAGCCACCCAC 3787
QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
Db 3788 CAGCAGTGTGCGCCAGAAAGCATCTATTCTCGATG-----GTTCAGGCTGGAGCA 3841
QY 1280 LysArg 1281
Db 3842 AAGCGC 3847

RESULT 10
US-09-450-105-1
; Sequence 1, Application US/09450105
; Patent No. 6169166
; GENERAL INFORMATION:
; APPLICANT: Kimberly Anne Brun
; APPLICANT: Richard James Chenery
; APPLICANT: Harma Ellens
; APPLICANT: John Anthony Feild
; APPLICANT: Lin Yue
; TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: ENCODING RAT MDRI2 AND SCREENING METHODS THEREOF
; FILE REFERENCE: GP-50008-DI
; CURRENT APPLICATION NUMBER: US/09/450,105
; EARLIER FILING DATE: 1999-11-29
; EARLIER APPLICATION NUMBER: 09/120,513
; EARLIER FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4233
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-450-105-1

Alignment Scores:
Pred. No.: 0 Length: 4233
Score: 5291.50 Matches: 1030
Percent Similarity: 90.48% Conservative: 130
Best Local Similarity: 80.34% Mismatches: 113
Query Match: 81.75% Indels: 9
DB: 4 Gaps: 5

US-09-672-725c-27 (1-1281) x US-09-450-105-1 (1-4233)
QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 26 ATGGAGTTGAAGAGGGCTTAACGGAACAGCAGCAAGAACTTCTCAAGATGGGCAA 85
QY 21 LysSerLysLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 40
Db 86 AAGAGTAAAG-----GAGAAGGAGAAGAACCTGCTGTGGCATATTCGGATGTTT 139
QY 41 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 60
Db 140 CGCTATGCAGATGGCTTGACAAGCTGTGCATGGCTCTGGGAACCTCTCGCTCATCATC 199
QY 61 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 80
Db 200 CACGGAACCCCTCTCCCTCCTCTGCTGTGTTGCGATACATGACAGATAGTTTACC 259
QY 81 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 100
Db 260 CAAGCA-----GAGACCCGCATCTGCCGAGCGTTACTAATCAAGTGAATCAAC 310

```

```

QY 101 AsnThrGlnHisPheIle---AsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 311 AGTACACAGACCGCTCAGCGACAGCAGCTCTGGAGGAGGACATGGCCATGTACGCCCTACTAT 370
QY 120 TyrSerGlyIleGlyAlaGlyValLeuAlaTyrIleGlnValSerPheTrpCys 139
Db 371 TACACGGGCAATGGTGGCGGTGTCTATCGTTCATCGTTCATCGAGTTTTCATCTTTGGTGC 430
QY 140 LeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArg 159
Db 431 CTGGCAGCTGGGAGACAAATACACAAGATTAGGCAGAAAGTTTTCATCGCCATCATGAAT 490
QY 160 GlnGluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAsp 179
Db 491 CAGGAGATAGGCTGTTTGCAGTGAATACGCTGGGGAGCTCAACACCCGCTCACAGAT 550
QY 180 AspValSerLysIleAsnGluGlyIleGlyAspLysIleGlyMetPhePheGlnSerIle 199
Db 551 GACGTCTCCAAATTAATGACGGAATTCGTGACAACTTGAATGTTCTTTTCAGTCCATA 610
QY 200 AlaThrPhePheThrGlyPheIleValGlyPheThrArgGlyTrpLysLeuThrLeuVal 219
Db 611 ACGACATTTTCAGCGGTTTTTATAATAGGATTTATAAGTGTGTGGAAGCTAACCCCTGTA 670
QY 220 IleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSer 239
Db 671 ATTTGGCGCTCAGCCCTCTTATTGGGTGTTCATCTGCCATGTGGGCAAGGTACTGACT 730
QY 240 SerPheThrAspLysGluLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluVal 259
Db 731 TCATTTACTAATAAGAACTCCAGCTTATCGGAAGCTGGAGCAGTTCGGAAGAGATC 790
QY 260 LeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyr 279
Db 791 TTAGCAGCCATCAGAACTGTGATTGCGTTTGGAGGACAAAAGAGAACTTGAAGGTAC 850
QY 280 AsnLysAsnLeuGluAlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIle 299
Db 851 AATAAATAATTAGAAAGAGCTTAAAGAGTTGGCATAAAGAAAGCCATCAGGCGCAACATT 910
QY 300 SerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGly 319
Db 911 TCCATAGTATTGCTACCTGTTGCTATGCGTCTTATGCACCTGGCATTCCTGGTATGG 970
QY 320 ThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSer 339
Db 971 ACCTCCTTGGTCTCTCAAAATGAATATTCATTGGACAAGTGTACCGTCTCTTCTCT 1030
QY 340 ValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsn 359
Db 1031 ATTTTATTTGGGACTTTTCAGTATTGGACATTTAGCCCCCAACATAGAACCTTTGCAAT 1090
QY 360 AlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSer 379
Db 1091 GCAAGAGGCGCAGCTATGAATCTTCAAGATAATTGATAATGAGCAAGCATTTGACAGC 1150
QY 380 TyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnVal 399
Db 1151 TTCTCAACCAAGGAGCACAAACCCAGACAGTATTAATGGGAAATTTGGAATTTAAAAATGTT 1210
QY 400 HisPheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysVal 419
Db 1211 TACTTCAACTACCATCAGCAAGTGAAGTTAAGATCTTGAAGGGCCCTCAACCTGGAAGGTG 1270
QY 420 GlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrVal 439
Db 1271 AAGAGGGGCGAGCGGTAGCCCTGTTGGCAACAGTGGCTGTGGGAAAGACCAACTGTC 1330
QY 440 GlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAsp 459
Db 1331 CAGCTGCTGAGAGGCTCTACGACCCCAATAGAGGGGAGGTGAGTATCGACGACGAGGAC 1390
QY 460 IleArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluPro 479

```


Db 1391 ATCAGGACCATCAATGTGAGGTATCTCGGGAATCATTTGGGGTGGTGTAGTCAGGAACCC 1450
Qy ValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMet 499
Db 1451 GTGCTGTTGGCACCACCATGTCGGAACATTCGCTATGGCCGAGAAACGTCACCATG 1510
Qy AspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuPro 519
Db 1511 GATGAGATAGAAAGCTGTCAAGGAACCAATGCCTATGACTTCATCATGAAACTGCC 1570
Qy 520 asnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGln 539
Db 1571 CACAATTTAACACCCCTGGTGGTGAGAGGGGCGACAGCTAGTGGGGGACAGAAACAG 1630
Qy 540 ArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAla 559
Db 1631 AGGATCGCATTTGCCCGGCCCTGTGCGCAACCCCAAGATCCTTTTGTGGATGAGGCC 1690
Qy 560 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 579
Db 1691 ACGTCAGCCTTGGACACAGAAAGCGGCGGTTCAGGCCGCTCTCGATAAGGCTAGA 1750
Qy 580 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 599
Db 1751 GAAGCGCGACCATTCGTGATAGTCACCGCTTGTCTACAGTCGCGCAATGCTGACGTC 1810
Qy 600 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 619
Db 1811 ATTGCTGGTGTGATGGTGTCTATTGTGGAGCAAGGAAATCATGAAGAGCTCATGAAA 1870
Qy 620 GluLysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeu 639
Db 1871 GAGAAGGCAATTTACTTCAAACTTGTACACACAGACTAGAGGAAATGAATGAACCA 1930
Qy 640 GluAsnAlaThrGlyGlySerLysSerGluSerAspAlaLeuGluLeuMetSerProLysAsp 659
Db 1931 GGAATTAATGCTTATGAATCCCAAGTGACACTGGTGGCTCTGAGTTGACTTTCAGAAAAA 1990
Qy 660 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 679
Db 1991 TCAAAATCTCTTTAATA---AGGAGATCAATTCGCAAGAGTATCCACAGAACACAGAC 2047
Qy 680 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 699
Db 2048 CAGAGAGAGACTTAGTTCGAAGAGGATGTGGATGAAGATGTGCTATGCTTCCCTTT 2107
Qy 700 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 719
Db 2108 TGGCAGATCCTAAAGCTAAATATTAGTAATGGCCCTATTATTAGTTGTGGGTACTTTGT 2167
Qy 720 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 739
Db 2168 GCTGTTATAAATGGGTGCATACCAACAGAGTGTGGCCATAGTGTTCGAAGATTGTAGG 2227
Qy 740 IlePheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerVal 759
Db 2228 GTTTTTTCAGACGACGACGACCATGAACCAACAAACGGAATGTAACTGTGTTTCCCTT 2287
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2288 CTCTTTCTGTCATGGGAATGATTTCTTTGTTACGTACTTCTTCAAGCTTTCACATTT 2347
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2348 GGCAAACTGGAGAGATCCTCACCAGGACTCCGATACATGCTTCAAAATCCATGCTG 2407
Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2408 CGACAGGATATAAGCTGGTTGATGACCAATAAAACACCACTGGCTCGCTGACTACACAG 2467
Qy 820 LeuAlaAsnAspAlaAlaGlnValLysGlyValaIleGlySerArgLeuAlaValIleThr 839

Db 2468 CTCGCTAGTCAGCCTTCTAATGTTAAAGGGGCTATGGGCTCCAGGCTTCTCTAGTTACC 2527
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2528 CAGAATGTAGCAACCTTGGCAGAGAAATATCTTATCTTATCTTATGCTTATGCTGGCAGCTT 2587
Qy 860 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2588 ACACCTTTTACTTGTAGTAATATTACCACCTATGCTTGGTGGAAATATTGAAATGAAA 2647
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2648 CTGTTGCTCGTCAAGCCTTGAAGGACAGAAAGACAGCTAGATCTCTGGAGATCGCT 2707
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2708 ACAGAGCAATTTGAAACTTCCGCACTGTTGCTCTTGTACTAGGAGAGCAAGTTTGA 2767
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2768 ACTATGATGCCAGAGCTTGCAGATACCATACAGAAATGCTTTGAAGAAAGCACACGTC 2827
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2828 TTTGGGATCACCTTCGCTTCCAGGCGCATGATTTATTTTCTTCTATGCTGCTGTTTC 2887
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 2888 CGGTTTCGGTGCCTACTTGTGGTGACGAGAACTCATGACGTTTGAATAATGTTATGTTGTA 2947
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 2948 TTTTCTCTGCTGTGCTTGTGGTCATGGCAGGGAATACCACTTCATTCGCTCGCTGAC 3007
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisIleMetIleIleGluLysSerProLeu 1019
Db 3008 TACCGGAGGCCAAGCTCAGCATCCACATCATCAGGATCATTTGAGAAATCCCGAG 3067
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3068 ATTGACAGCTACAGCAGGAGGCTTGAAGCCTAATTTGGTTAGAAAGAAATGTGAATTT 3127
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3128 AATGAGTCATGTTCATCATCCACCGACCCCAACATCCAGTGCCTCAGGACTCAGC 3187
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3188 TTCGAGGTTGAAGAGGGGCAACGCTTCGCTGGTGGCAGCAGTGGCTGCGGGAAGAGT 3247
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1099
Db 3248 ACAGTGGTCCAGCTGCTCGAGCGCTTCTACAAACCCCATGGCTGGACAGAGTTCATAGAT 3307
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3308 GCCTAAGAAATAAACAACACTCAAGTCCAGTGGCTCCGCGCCACCTGGGCATTTGTGTC 3367
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3368 CAGGAGCCCATCTCTTTGACTGCAGCATCACCGAGAAACATCGCCTACGGAGAACAGC 3427
Qy 1140 ArgValValSerHisGluLeuIleValGlnAlaLysGluAlaAsnIleHisPhe 1159
Db 3428 CGTGTGCTGTCTCATGAGAGATCGTAGGCCCGCCAGGAGGCCAACATCCACAGTTC 3487
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3488 ATCGACTCACCTGCCCTGAGAAATACAAACACAGAGTGGGAGACAAAGGAGCTCAGCTGTCG 3547
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3548 GCGGGGAGAAAGCAGCGCATCGCCATCGCGCGCCCTCGTCGACAGACGCTCACATCTTA 3607

QY 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
 Db 3608 CTCTGGATGAGCGCATCAGCTCTGGATACGAGAGTGAAGAAGTGGTCCAGGAAGCG 3667
 QY 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
 Db 3668 CTGGACAAAGCCAGGAAGCGCACCTGCTGATGTCGCGCACCGCTGTCCACCATC 3727
 QY 1240 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
 Db 3728 CAGAACCGACACTGATCGTGGTGATTCAGAACGGCCAGGTCAAGGAGCAGCGCACCCAC 3787
 QY 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla 1279
 Db 3788 CAGCAGCTGTGCGCCAGAAAGCATCTATTTCTCGATG-----GTTACGGCTGGAGCA 3841
 QY 1280 LysArg 1281
 Db 3842 AAGCGC 3847

RESULT 11
 US-08-461-823-1
 ; Sequence 1, Application US/08461823
 ; Patent No. 5593840
 ; GENERAL INFORMATION:
 ; APPLICANT: Bhatnagar, Satish K.
 ; APPLICANT: George Jr., Albert L.
 ; APPLICANT: Nazarenko, Irina
 ; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACID SEQUENCES
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OncorPharm, Inc.
 ; STREET: 200 Perry Parkway
 ; CITY: Gaithersburg
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20877

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/461,823
 FILING DATE: 05-JUN-1995
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/168,621
 FILING DATE: 16-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/010,433
 FILING DATE: 27-JAN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Karta, Glenn E.
 REGISTRATION NUMBER: 30,649
 REFERENCE/DOCKET NUMBER: PA-0012 CIP 2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301 527-2058
 TELEFAX: 301 208-6997

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2726 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-461-823-1

Alignment Scores: 0 Length: 2726
 Pred. No.:

Score: 3631.00 Matches: 717
 Percent Similarity: 96.28% Conservative: 34
 Best Local Similarity: 91.92% Mismatches: 29
 Query Match: 56.09% Indels: 0
 DB: 1 Gaps: 0

US-09-672-725C-27 (1-1281) x US-08-461-823-1 (1-2726)

QY 502 IleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsnLys 521
 Db 2 ATTGAAAGCTGTCAAGGAGCAATGCTATGACTTTATCATGAAATGCTCATAAA 61
 QY 522 PheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGlyGlnLysGlnArgIle 541
 Db 62 TTGTGACACCTCGTTGGAGAGAGAGGGCCAGTTGAGTGGTGGCAGAAAGCAGAGGATC 121
 QY 542 AlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThrSer 561
 Db 122 GCCATTGCACGTGCCCTGGTTCGCAACCCCAAGATCCTCTGCTGATGAGCCACGTCA 181
 QY 562 AlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLysGly 581
 Db 182 GCCTTTGGACACAGAAAGCGAAGCAGTGGTTTCAGGTGGCTCTGGATAAGCCAGAAAGGT 241
 QY 582 ArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla 601
 Db 242 CGACCCACCATTTGATAGCTCATCGTTGTCTACAGTTCGTAAATGCTGAGCATCGCT 301
 QY 602 GlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGluLys 621
 Db 302 GGTTCGATGATGGAGTCATTTGGGAGAAAGAAATCATGATGAATCATGAAGAGAAA 361
 QY 622 GlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGluAsn 641
 Db 362 GGCATTTACTTCAAACTTGTCCAAATGCAGACAGCAGGAAATGAAGTTGAATTAATAAT 421
 QY 642 AlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSerGly 661
 Db 422 GCAGCTGATGAATCCAAAAGTGAATGATGCCCTTGGAAATGCTCTCAATGATTCAGA 481
 QY 662 SerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAsp 681
 Db 482 TCCAGTCTAATAAGAAAGATCAACTCGTAGAGTGTCCGTGGATCACAAGCCCAAGAC 541
 QY 682 ArgLysLeuGlyThrLysGluAspLeuAsnValProProValSerPheTrpArg 701
 Db 542 AGAAAGCTTAGTACCAAGAGGCTCTGGATGAAATATACCTCCAGTTCTCTTTGGAGG 601
 QY 702 IleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAlaIle 721
 Db 602 ATTATGAAGCTAAATTTAACTGAATGGCTTATTTTGTGTGGTGATTTTGTGCCATT 661
 QY 722 IleAsnGlyGlyLeuGlnProAlaPheSerIlePheSerArgIleIleGlyIlePhe 741
 Db 662 ATAAATGGAGGCTCGCAACCAAGCATTTTCAATAATATTTTCAAGATTAATAGGGGTTTT 721
 QY 742 ThrArgaspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe 761
 Db 722 ACAGATTTGATGATCTCGAACAACCAAGCAGAGATAGTAACTTTTTCACATATTTT 781
 QY 762 LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys 781
 Db 782 CTAGCCCTTGGATTAATTTCTTTTATTTACATTTTCTTCAGGGTTTTCACATTTGGCAA 841
 QY 782 AlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGln 801
 Db 842 GCTGGAGAGATCCTCAGCAAGCGGCTCCGATACATGTTTTCATGATCCATGCTCAGAC 901
 QY 802 AspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla 821
 Db 902 GATGTGAGTTGGTTTGTATGACCCCTAAACCAACCACTGGAGCATTTGATCAGCGCTGCC 961
 QY 822 AsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn 841

Db 962 AATGATGCTGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGCTGTAATTAATACCCAGAA 1021
Qy 842 lIeAlaAsnLeuGlyThrGlyllelleSerLeuIlefyGlyTrpGlnLeuThrieu 861
Db 1022 ATAGCAAACTTGGGACAGGAATAATATATCTTCACTATGCTGTGGCAACTAACACTG 1081
Qy 862 LeuLeuLeuAlaIleValProIlelleAlaIleAlaGlyValValGluMetLysMetLeu 881
Db 1082 TTACTCTTTAGCAATTTGATCCCATCATTCGAATGACAGAGTGTGGAATGAAATGTTG 1141
Qy 882 SerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAlaThrGlu 901
Db 1142 TCTGGCAAGCACTGAAGATGAAGAACTAGAGGTGCTGGGAAGATGCTACTGAA 1201
Qy 902 AlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyMet 921
Db 1202 GCAATAGAAACTTCCGAACCGTGTCTTCTGACTCAGGACGAGAGTTTGAACATATG 1261
Qy 922 TyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGly 941
Db 1262 TATGCTCAGAGTTGAGGTACCATACAGAACTCTTTGAGGAAGCACACATCTTTGGA 1321
Qy 942 ValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPhe 961
Db 1322 ATTACATTTTCTTCACCCAGCAATGATGATATTTTCTATGCTGGATGTTCCGGTT 1381
Qy 962 GlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSer 981
Db 1382 GGAGCCTACTTGGTGCAATAACTATGAGCTTTGAGGATGTTCTGTTAGTATTTCA 1441
Qy 982 AlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyrAla 1001
Db 1442 GCTGTGCTTTGGTGCCATGCGCGTGGGCAAGTCAGTTCAATTTGCTCCTGACTATGCC 1501
Qy 1002 LysAlaLysValSerAlaAlaHisValIleMetIlelleGluLysSerProLeuIleAsp 1021
Db 1502 AAAGCCAAAATATCAGCAGCCACATCATCATGATCATTTGAAAAACCCCTTGATTGAC 1561
Qy 1022 SerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGlu 1041
Db 1562 AGCTACAGCAGGAGGCGCTTAATCGGAACACATTTGGAAGAAATGTCACTTTGGTGA 1621
Qy 1042 ValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeuGlu 1061
Db 1622 GTTGATTTCACTATCCACCCGACCGGACATCCAGTGCTTCAGGACTGAGCCTGGAG 1681
Qy 1062 ValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrVal 1081
Db 1682 GTGAGAAGGGCCAGACGCTGCTGTGGGACAGTGGCTGTGGGAAGAGCACAGTG 1741
Qy 1082 ValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGlyLys 1101
Db 1742 GTCACGCTCCTGAGCGGTTCTACAGCCCTTGGCAGGGAAGTCTGCTGATGGCAA 1801
Qy 1102 GluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGlu 1121
Db 1802 GAAATAAGCGACTGAATGTTAGTGGCTCCGAGCACACCTGGGCATCGTGTCCAGGAG 1861
Qy 1122 ProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgVal 1141
Db 1862 CCCATCCTGTTGACTGACATTTGCTGAGAACATTTGCTATGGAGACAAACGCGGGTG 1921
Qy 1142 ValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisPheIleGlu 1161
Db 1922 GTGTCAGAGAAAGATGCTGAGGCGAGCAAGAGGCCAACATACATGCTTCATCGAG 1981
Qy 1162 ThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGly 1181
Db 1982 TCAGTGCCTAATAATATAGCACTAAAGTAGGAGACAAAGGAACCTCAGCTCTGTGGTC 2041
Qy 1182 GlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeu 1201

Db 2042 CAGAAACAACGCAATGCCATAGCTCGTCCCTTGTAGACAGCCTCATATTTGCTTTTG 2101
Qy 1202 AspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeuAsp 1221
Db 2102 GATGAGCCACGTCAGCTCTGGATGATCAGAAAGTGAAGGTTGTCCCAAGAGCCCTGGAC 2161
Qy 1222 LysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsn 1241
Db 2162 AAAGCCAGAGAAGCGCACCTGCATTTGATTTGCTCACCGCTGTCCACCATCCAGAA 2221
Qy 1242 AlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGlnGln 1261
Db 2222 GCAGACTTAATAGTGTGTTTCAGAAATGCGAGATGCAAGAGCATGCCAGCATCAGCAG 2281
Qy 1262 LeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAlaLysArg 1281
Db 2282 CTGCTGCACAGAAGGCATCTATTTTCAATGTCAGTGTCCAGCTGGAACAAAGCGC 2341
RESULT 12
US-08-612-734B-1
; Sequence 1, Application US/08612734B
; Patent No. 5914246
; GENERAL INFORMATION:
; APPLICANT: Peery, Robert B.
; APPLICANT: Skatrud, Paul L.
; APPLICANT: Tobin, Matthew B.
; TITLE OF INVENTION: Multiple Drug Resistance Gene of
; TITLE OF INVENTION: Aspergillus Fumigatus
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center, DC1501
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.A.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,734B
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Craig, Anne I.
; REGISTRATION NUMBER: 32,976
; REFERENCE/DOCKET NUMBER: X-9681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-354-9570
; TELEFAX: 617-354-4043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4047 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4047
US-08-612-734B-1

Alignment Scores:
Pred. No.: 3,51e-276 Length: 4047
Score: 2514.00 Matches: 559
Percent Similarity: 59.35% Conservative: 228
Best Local Similarity: 42.16% Mismatches: 465
Query Match: 38.84% Indels: 74
DB: 2 Gaps: 17
US-09-672-725C-27 (1-1281) x US-08-612-734B-1 (1-4047)

QY 1 MetAspProGluGlyArgLysGlySerAlaGluLys----- 13
Db 148 ATGGCGCTCCACATGGCAGAAAGGATCAGGGAAAGCTGTGCACCTGAACGATGAT 207
QY 14 -----AsnPheTrpLysMetGlyLysLysSerLysLysLysGluLysLysGlu 29
Db 208 TCTCTCTTTCCCATCTCCAGGAGCAGCAGAGAGAGTACTCAAGCAGCAGCTTGATGCA 267
QY 30 LysLysProThrValSerThrPheAlaMetPheArgTyrSerAsnTrpLeuAspArgLeu 49
Db 268 CCTCTGTGAAGTCTCTTCTTCACTTGTATCGCTATCGCTCGAGAAAGCAGCATCCGTG 327
QY 50 TyrMetLeuValGlyThrMetAlaAlaIleIleHisGlyAlaAlaLeuProLeuMetMet 69
Db 328 ATCATACTCGTAGCCCACTGTGCTATTGTGCTGGTGGCGCTCTTCCCTTGTTTCA 387
QY 70 LeuValPheGlyAsnMetThrAspSerPheAlaAsnAlaGlyIleSerArgAsnLysThr 89
Db 388 ATTCTCTCGGTTCACATAGCCTCAGCGTTC-----CAGGGCATATCTCTGGGT---ACT 438
QY 90 PheProValIleIleAsnGluSerIleThrAsnAsnThrGlnHisPheIleAsnHisLeu 109
Db 439 ATGCCCC-----TATCATGAGTTCATCAC---- 462
QY 110 GluGluGluMetThrThrTyrAlaTyrTyrSerGlyIleGlyAlaGlyValLeuVal 129
Db 463 -----AACTGACTAAGATGTGTTTACTTGTATCTCGGTATGCCAGTGTGTC 516
QY 130 AlaAlaTyrIleGlnValSerPheTrpCysLeuAlaAlaGlyArgGlnIleLeuLysIle 149
Db 517 ACAGTCTATGTGAGCAGCCGTGGTTCATTATATCTGCGCAACATCTCACACAGAAGATC 576
QY 150 ArgLysGlnPheHisAlaIleMetArgGlnGluIleGlyTrpPheAspValHisasp 169
Db 577 CGTGAATAATTCTTGAGGCTATCTTGAGGAGCAATATGGCTTACTTCGACAAGTTGGGC 636
QY 170 ValGlyGluLeuAsnThrArgLeuThrAspValSerLysIleAsnGluGlyIleGly 189
Db 637 GCCGGTGAAGTTACCACGCGTATCACTGCTGATACCAACCTCAGCAGCCCATCTCT 696
QY 190 AspLysIleGlyMetPhePheGlnSerIleAlaThrPheThrGlyPheIleValGly 209
Db 697 GAGAAAGTTGGTCTCATTCTGACCCATTCGCCACATTTGTAACCGCATTTATTGTGCC 756
QY 210 PheThrArgGlyTrpLysLeuThrLeuVal-----IleLeuAlaIleSerPro 225
Db 757 TACGTCAAGTATGGAAGTTGGCTCTGATCTGATCTACCTCAACTATCGTCGCGCTGTCATG 816
QY 226 ValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPheThrAspLysGlu 245
Db 817 GTTATGGGAGGTGGG-----TCGAGGTTTATTGTGAAGTACAGCAAGAAATCT 864
QY 246 LeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeuAlaAlaIleArgThr 265
Db 865 APTGAAAGTTATGTGCTGGTGGTGAAGTCTCGCGAAGAGTATCATCAGTCCATTCGGAAT 924
QY 266 ValIleAlaPheGlyGlnLysLysGluLeuGluArgTyrAsnLysAsnLeuGluGlu 285
Db 925 GCTACCGGTTTCGGCACTCAGATGAAGTCCGCAAGCAANTACGAAACCATCTGGCTGAG 984
QY 286 AlaLysGlyIleGlyIleLysLysAlaIleThrAlaAsnIleSerIleGlyAlaAlaPhe 305
Db 985 GCTGAAAAATGGGCGTCAACACAGGTCATCTTGTATGATGATGATGGTGGTATGTTTC 1044
QY 306 LeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSerLeuValLeuSer 325
Db 1045 GGTATCATGTTCTCGAATATGGTCTCGGTTTCTGGATGGGATCTCGCTTCTGTTGGT 1104
QY 326 SerGluTyrThrIleGlyGlnValLeuThrValPhePheSerValLeuIleGlyAlaPhe 345
Db 1105 AAAGAGTCAACGTTGGCCAAAGTTCTGACAGTTTGTATGATCTATCTGATCGGTTCGTTTC 1164

QY 346 SerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArgGlyAlaAlaTyr 365
Db 1165 AGTTTGGCAAGCTGCGCCCAATGGTCTAGGCTTTAGCAATGGTGTCTGCGCGCGC 1224
QY 366 GluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSerLysSerGlyHis 385
Db 1225 AGATTTCAGCAGCATGATGCGCAGATGCGCAGTGGCCCTTATTCGAGGAAGGAAG 1284
QY 386 LysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPheSerTyrProSer 405
Db 1285 GTACTCGACCACTTTGAGGAATATCGAATTTGCAATGTCAACACATCTACCTTCA 1344
QY 406 ArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSerGlyGlnThrVal 425
Db 1345 AGACCCGAAGTTACAGTATCGAAGATGTCTTTATCGATGCGCGCGGGAAGACTACC 1404
QY 426 AlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeuMetGlnArgLeu 445
Db 1405 GCATTGGTGGCCCATCTGGCTCTGGAAGAGTACTGTGTGCGCTTGGTGAGCGGTTT 1464
QY 446 TyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArgThrIleAsnVal 465
Db 1465 TACCTTCAGTAGGAGCGCAGGTATGCTGGACGCCATGATATCCAAACCTCAACCTC 1524
QY 466 ArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAlaThrThr 485
Db 1525 CGTTGGCTGCGCAGCAGATCTCTTGTGAGCCAGGAACCTGTTCTTTTCAGCACCACG 1584
QY 486 IleAlaGluAsnIleArgTyrGlyArg-----GluAsnValThrMetAsp 500
Db 1585 ATCTTAGAATCATCGACATGGCTGTGTCGCCACCAATTCGAGCATGAGTCGAAGGAC 1644
QY 501 GluIle-----GluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMet 516
Db 1645 AAGATTAGAGAGCTCGTTGAGAATGCGGCCAAGATGCGCAATGCTCATGATTTATTATG 1704
QY 517 LysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSerGlyGly 536
Db 1705 GCTCTGCTGAAGTTACGATACCAATGTGGGTGAGCGTGTCTTCTTACTTTTCAGGAGT 1764
QY 537 GlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeu 556
Db 1765 CAGAAGCAACGATTATGCCATTGCTGCGGCCATGTCTAGTCACCCCAAGATCTGTGCTT 1824
QY 557 AspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAsp 576
Db 1825 GATGAAGCTACATCAGCTTTGGATACCAAGTCCGAGGCGCTGCTCCAAAGCGCTCTGTAT 1884
QY 577 LysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsn 596
Db 1885 AAAGTCCGAGGAGTAGACTACCATTTGTCATTGCTCACCCTTTGTCAACAAATCAAAACA 1944
QY 597 AlaAspValIleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGlu 616
Db 1945 GCCCACAACATTTGTGCCATGCTCGCGGAAAGATTTCCGCAACAGGAGACACATGAGAA 2004
QY 617 LeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGlu 636
Db 2005 TTGGTCGATCGCAAGGACGCTACTATAACTTGTGGAGCGCAAGTATCAACAGGAGGAG 2064
QY 637 IleGluLeuGluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSer 656
Db 2065 AAGAGCAGCAA-----GCTCTGGAAGCCGACGCC---GACATGGAC 2103
QY 657 ProLysAspSerGlySerSerLeuIleLysArg-----ArgSerThrArgArg 672
Db 2104 GCCGATGATTTTGTCAAGAGGGGTTACTCGCATCAAGACTCAGCTTAGCAGCTCGAAT 2163
QY 673 SerIleHisAlaProGlnGlyGlnAspArg-----LysLeuGlyThrLysGlu 688
Db 2164 TCTCTCGATGCTAGATGAGAGCGCGCTTGGAGATGAAGCGACCGACGCAAAAAA 2223
QY 689 Asp-----LeuAsnGluAsnValPro-----ProValSerPhe 699

Thu Nov 7 09:13:24 2002

Db 1795 CTCCTGGATGAAGCTACTTCGGCCCTTGGACACAAAAATCCGAAGCGCTGTTCAAGCAGCT 1854
Qy 575 LeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrVal 594
Db 1855 TTGGAGAGGCGAGCTGAGGCGGAACACTATATTGTGATCCCTCATCGCTTCCACGATC 1914
Qy 595 ArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHis 614
Db 1915 AAAACGGCGCACACATTTGCTGCTCAATGCGAAATTTGCTGAACAAGAACTCAC 1974
Qy 615 AspGluLeuMetLysGluLysGlyIleThrPheLysLysValThrMetGlnThrArgGly 634
Db 1975 GATGAATTTGTTGACCGGAGCGCTTATTCGCAAACTTTGGAGGCTCAACGTTATCAAT 2034
Qy 635 AsnGluIleGlu-----LeuGluAsnAlaThrGlyGluSerLysSerGluSerAsp 651
Db 2035 GAACAGAAGAGCTGACCGCTTGGAGACGCCGACGCTGAGGATCTCAGCAATGCAGAT 2094
Qy 652 AlaLeuGluMetSerProLysAspSerGlySerSer-----Leu 664
Db 2095 ATTGCCAAAATCAAAACTCGTCAAGCGCATCATCCGATCTCGACGGGAAAACCCACAAACC 2154
Qy 665 IleLysArgArgSerThrArgSerIleHisAlaProGlnGlyGlnAspArgLysLeu 684
Db 2155 ATTGACCGCAGCGGCHCCACAAAGCTGTTCACGCGGATCTTCTTAAAGACCC--- 2211
Qy 685 GlyThrLysGluAspLeuAsnGluAsnValProProValSerPheThrPArgIleLeuLys 704
Db 2212 -----CCCGAACAACCTCCGAAATACTATTATGAGCGCTGCTCANA 2253
Qy 705 -----LeuAsnSerThrGluTrpProThrPheValValGlyIlePheCysAla 720
Db 2254 TTTTGTGCTTCCCTTCACCGCCCGTAATCCCGATACGCTCATCGCTTGTCTCTCA 2313
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIlePheSerArgIleIleGlyIle 740
Db 2314 GTGTAGCTGGTGGTGGCCCAACCCACGACAGTGTCTATATGCTAAAGCCATCAGCAC 2373
Qy 741 PheThrArgAspGluAspProGluThrLys--ArgGluAsnSerAsnMetPheSerVal 759
Db 2374 CTCCTCGCTCCAGAAATCACAATATAGCAAGCTTCGACATGATGCGGATTTCTGGTCA 2433
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2434 ARGTTCTCGTGGTGGTATCATCTCACTTATCAGCAGTCAACCAATGGTGTGCTCATTT 2493
Qy 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2494 GCCGTATGCTCCGAGAGACTTATTCTGCGCGGAGAGACACTGCCCTTCGGAGATATC 2553
Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2554 CGTCAAGACATGCTTTCTTTGACAAAGGAGAGAATAGCACCGCGGCTCTGACCTCTTC 2613
Qy 820 LeuAlaAsnAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2614 CTGTCCACCGACGAAAGCATCTCTCCGGTGTAGCGGTGATCTAGGCACGATCTTG 2673
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2674 ATGACCTCCACGACCTTAGGACGGCTATCATTTATGCGCTGGCGATTTGGTGAATG 2733
Qy 860 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 879
Db 2734 GCCTTAGTTGATCTCGGTTGCGGTTCTCTGCGATCGGCTTCTACCGATTCAT 2793
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2794 ATGCTAGCCAGTTTCAATACGCTCCCAAGCTTGCTTATGAGGATCTGCAAACTTTGCT 2853
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2854 TGGAGGCTACATCGTCTATCCGACAGTGTGCTATTAACCCGGGAAAGGATGTCTGG 2913

Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2914 GAGATTACCATGCCAGCTTGAGCACAAGGACAGGACCATTAATCTCTCTCTGTGAGG 2973
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2974 TCATCCCTGTTATATGCTGTCGACGACCTTGTCTTCTTCGCTGGCTCGGCTGGTTT 3033
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal 979
Db 3034 TGGTACGAGGAGGACACTTCTTGGTCACACGATATGACATTTTCCGCTCTCTTTGTTGT 3093
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlnValSerSerPheAlaProAsp 999
Db 3094 TTCCTCGAGATCTCTTGTGCTCAATCCGCGGCGACCTTTTCTTCTTGCCACGAC 3153
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3154 ATGGCAAGGCGGAAGATGCGCGCCCGCAATCCGACGACTGTGTCACCGCAAGCCACAA 3213
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3214 ATTGATACTGCTCTGAAGAGGCGGAGAGCTCGAAACGGTGGAGGTGAATCGAATTT 3273
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3274 AGAAGCTGCACCTCAGATACCCGACCCGCCAGAACAGCCTGCTCTCGCGGCTTGAC 3333
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3334 CTGACCGTGAAGCTCGACAATATGTCGCTGTGCGACCCGCGGTGTGGCAAGGT 3393
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValIleAsp 1099
Db 3394 ACCACCATTCATTTGCTGAGCGCTTTTACGATGCGATTCGCGGCTCCATCTCTGTTGAT 3453
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3454 GGAAGACATAGTAACATAATATCAACTCTACCGACGCTTCTGTCATGTGTCAGC 3513
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3514 CAGGAGCGCACTGTACCAGGCGCACCATCAGAAACATCTTACTTGGTATTGTCGAA 3573
Qy 1140 ArgValValSerHisGluIleValGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
Db 3574 GATGACGTACCGGAGAAATTTGATTAGGCTTGAAGGACGCTAATATCTACGACTTC 3633
Qy 1160 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1179
Db 3634 ATCATGTCGCTCCCGGAGGCTTTAATACAGTTGTTGGCAGCAAGGAGGCGCATGTTGCT 3693
Qy 1180 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1199
Db 3694 GCGGCCCAAGAACAGCTGTGCGCATTTGCCGAGCCCTTCTTCGGGATCCCAAAATCCTT 3753
Qy 1200 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1219
Db 3754 CTTCTCATGAAGCGACGCTCAGCCCTCAGCTCCGAGTCCAGAAAGGCTGTCAGGCGGCT 3813
Qy 1220 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1239
Db 3814 TTGGGATCGCGTGGCCGAGGCGGAACACATCGCCGTGTCACACCGACTCAGCAGCAT 3873
Qy 1240 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1259
Db 3874 CAAAGGCGGACGTTATCTATCTTTTCGACCAAGCAGATCGTCGAAAGCGGAAACGAC 3933
Qy 1260 GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaCylala 1279
Db 3934 AGCGAACTGCTCCAGAAAAGGCGGCTACTACGAGCTGGTCAACTTGCAGAGCTTGGC 3993

Db 1435 CGAUCUACAUCCGCUUUGCGGUAUGCUUUUUGUGGAGCCCAUGACAUCACGAGCCUC 1494
Qy 464 AsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAla 483
Db 1495 AAUCCGCGGCUUGCGCAACAGAUUCUUGGUUAGCCAGGAGCCUUGUCUUUUUGGC 1554
Qy 484 ThrThrIleAlaGluAsnIleArgTyrGly-----ArgGluAsnValThr 498
Db 1555 ACCAGGAUUUAAGAAUUAUAGGCACGCGUCUACUGCGCAACAAAGAGGAUAGAUCC 1614
Qy 499 MetAspGlu-----IleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 514
Db 1615 GAGGAUAGGUCCGGAACUACUACGAGAACCGCGCAAAAUAGGCGAAUGCUACUAGACUUU 1674
Qy 515 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSer 534
Db 1675 AUUACUGCCUUGCGUAAGUUAUGAGACCAAAUUGUGGCGAGCGUGGUUCUCCUUUCA 1734
Qy 535 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu 554
Db 1735 GUGGCGCAAAACAGCGCAUUGCAUUGCGCGCGGUUGUUGAGUGACCCCAAAAUCCUG 1794
Qy 555 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla 574
Db 1795 CUCCUGGAUGACUACUUGCGCUUGGAGACAAAUUCCGAAAGGCGUGUUAAGCAGCU 1854
Qy 575 LeuAspLysAlaArgLysGlyArgThrIleValIleAlaHisArgLeuSerThrVal 594
Db 1855 UUGGAGAGCGCAGCUAGAGCGCAACUACUUAUUGUGAUGCGUACUGCCUUUCCAGAU 1914
Qy 595 ArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHis 614
Db 1915 AAAACGGCGACAAACAUUGGUUGUUGCAUUGGCAAAAUUGGCAAAUUGGCAAGACUCAC 1974
Qy 615 AspGluLeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGly 634
Db 1975 GAUGAAUUGUGUAGCCGCGCGCGCUUAUUGCAAAUUGUGGAGCGCUCACGUAUCAA 2034
Qy 635 AsnGluIleGlu-----LeuGluAsnAlaThrGlyGluSerLysSerGluSerAsp 651
Db 2035 GAACAGAAGGAAGCUGACCGCUUGGAGGACCGCGCGUAGGAGUACUACGAUAGCAGAU 2094
Qy 652 AlaLeuLeuMetSerProLysAspSerGlySerSer-----Leu 664
Db 2095 AUUGCCAAAUACAACUAGCGUACGCAUCAUCAUCGACGCAAAACCCCAAC 2154
Qy 665 IleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAspArgLysLeu 684
Db 2155 AUUGACGCGACGGGACCCCAACAGUCUUGUUUCCAGCGCAUUCUUAAGACCC--- 2211
Qy 685 GlyThrLysGluAspLeuAsnGluAsnValProValSerPheThrPargIleLeuLys 704
Db 2212 -----CCGAAACACUCCGCAAAUACUUAUUGAGCGGCUUGCUCAAA 2253
Qy 705 -----LeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db 2254 UUGUUGCUUCCUACAACCGCCUAGAAUCCGUAUGCUACUGGCUUGUCUUCUCA 2313
Qy 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2314 GUGUUGAGUGGUGGUGGCGCAACCCCAAGCAGUGCUUAUUGCUAAAGCCCAUCACACA 2373
Qy 741 PheThrArgAspGluAspProGluThrLys---ArgGlnAsnSerAsnMetPheSerVal 759
Db 2374 CUCCUGCUCCAGAAACACAAUUAAGCAAGCUUGCAGAUAGUGGGAUUCUGGUCAUUG 2433
Qy 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2434 AUGUUCUUGGUGGUGGUAUCAUGUUAUACACGCAAGUACCCAAACCAUUGGUGGUAUU 2493
Qy 780 GlyLysAlaGlyIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2494 CCCGUAGUCCGAGAGACUUAUUGCGCGCGAGAGACGACUGCCUUCGCGAGUAUCUC 2553

Qy 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2554 CGUCAAGACUUGCUUUCUUGGACAGGAGAAUAGCACCGCGCGCUGACCUUUC 2613
Qy 820 LeuAlaAsnAspAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2614 CUGUCCACCGAGAGCAAGCAUCUCGCGGUGUAGGUGUGACUCUAGCGCAGACUUG 2673
Qy 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2674 AUGACCUCCAGCAGCCUAGGAGCGGUACUUAUUCUUGCGCGGAGUUGGUGAAUUG 2733
Qy 860 ThrLeuLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLys 879
Db 2734 GCCUUAUUGUAUCUGGUGUGCGGUUCUGCGAUGCGGUCUUCACCGAUUCUAU 2793
Qy 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2794 AUGUAGCCCGAUUUCAACACGCUCCAAAGCUUUAUAGGAGUUCUGCAAAUUGCU 2853
Qy 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2854 UGCAGGCUACUUGCUUACCGCACAGUUGCGUCAUUAACCGGGAAGGAGUUGUG 2913
Qy 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2914 GAGAUUACCAAGCCAGCUUGACGACAAAGCGCAGCAGCAGUUAUCUCUGUUGAGG 2973
Qy 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2974 UCAUCCUGUAUUGCGUGCGCAGCAGCUUUGUUUUCUUGCGUGGCGUGGCUU 3033
Qy 960 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 3034 UGUACGAGGAGCACUUCUUGGACACGAGUACAGAUUUUCCGCUUUCUUGUUGU 3093
Qy 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3094 UUCUCCGAGAUUCUUCUUGGUCUAAUCCGCGGCGCACCUCUUCUUCUUGCCACACAG 3153
Qy 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019
Db 3154 AUGGCAAGCGAGAAUGCGCGCGCGCAUUCGACGACUGUUCGACCGAAAGCCCAA 3213
Qy 1020 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1039
Db 3214 AUUGAAUACUGUGGAGAGGCGAGAGCUCGAAAGCUCGAAAGGUGGAGAAUUG 3273
Qy 1040 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1059
Db 3274 AGGAACGUGUACUUCAGAUACCGGACCGCCAGAACAGCCUUGUCCUGCGCGUUGGAC 3333
Qy 1060 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1079
Db 3334 CUGACCGUGAGAGCCUGACAAUUGUGCGUUGCGGACCGAGGUGGAGGAGAAUUG 3393
Qy 1080 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuLeAsp 1099
Db 3394 ACCACCAUUGCAUUGCGUGAGCGUUUACGAGUUGCGGUGGUGGUGUUGUUGAU 3453
Qy 1100 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1119
Db 3454 GGAAGGACAAUAGUAAACUAAAUACUACUCCGACGAGCUUUCUGACUGGUCAGC 3513
Qy 1120 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1139
Db 3514 CAGGAGCCGACACUGUACCGAGGCGCACAUCAAGAGAAAUCAUUCUUGGUUUGUCAA 3573
Qy 1140 ArgValValSerHisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisPhe 1159
Db 3574 GAUGACGUACCGGAGAGAUUUCUUGAUUAGGCGUUGCAAGGAGCUUAUUAUCAGCACU 3633

QY 304 AlaPheLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrGlyThrSerLeuVal 323
Db 1015 ARGTTTGGCTTATGTAACCTACGCTTCTGGCTTCTGGATGGGTTCGTTTCCGTG 1074
QY 324 LeuSerSerGluTyrThrIleGlyGlnValLeuValPhePheSerValIleIleGly 343
Db 1075 GTAGATGGTGCAGTCGATGGGTGATATTCTACAGTTCTCATGGCCATCTTGATCGGA 1134
QY 344 AlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaAargGlyAla 363
Db 1135 TCGTTCTCTCTGGGAACGTTAGTCCAAATGCTCAAGCATTTACAAACGCTGTGGCCGCG 1194
QY 364 AlaTyrGluIlePhePheIleLeuAsnLysProSerIleAspSerTyrSerLysSer 383
Db 1195 GCGCAAGATATTGGGAACGATCGATCGCCAGTCGCCATTAGATCCCATATTCCGAACGAA 1254
QY 384 GlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPheSerTyr 403
Db 1255 GGAAGACGCTCGACATTTTGGGGCCACATTGATGACCAATGTCAAGCATATTTAC 1314
QY 404 ProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSerGlyGln 423
Db 1315 CCATCTAGACCCGAGTCCACCTGATGGAGGATGTTCTCTGTCATGCCCGCTGGAAA 1374
QY 424 ThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeuMetGln 443
Db 1375 ACAACGCTTGTAGTGGCCCCCTCTGGCTCTGGAAGAAGTACGGTGGTGGTTGAG 1434
QY 444 ArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArgThrIle 463
Db 1435 CGATTCTACATGCTGTTCGCGGTACGGTTTGTCTGGATGGCATGACATCAAGGACCTC 1494
QY 464 AsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAla 483
Db 1495 AATCTCGTGGCTTCGCCAACAGATCTCTTGGTTAGCCAGAGCCTGCTCTTTTGGC 1554
QY 484 ThrThrIleAlaGluAsnIleArgTyrGly-----ArgGluAsnValThr 498
Db 1555 ACACGATTTATAGAAATATTAGCAGCGTCTCATCGGCACAAAGTACGAGAATGAATCC 1614
QY 499 MetAspGlu-----IleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 514
Db 1615 GAGGATAAGTCCGGAACTCATCGAGAAGCGGCAAAATGGCGAATGCTATGACTTT 1674
QY 515 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaGlnLeuSer 534
Db 1675 ATTAATGCTTGGCTGAGGTATGAGACCAATGTTGGGCGGTGCTTCTCCTTTCA 1734
QY 535 GlyGlyGlnLysGlnArgIleAlaIleAlaAargAlaLeuValArgAsnProLysIleLeu 554
Db 1735 GGTGGCAGAAACAGCGCATTCGAATCGCCGTCGCGTTGTTAGTACCCCAAAATCCTG 1794
QY 555 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla 574
Db 1795 CTCCTGATGAAGCTACTTCGGCTTGACACAAATTCGAAGCGGTGTTCAAGCAGCT 1854
QY 575 LeuAspLysAlaArgLysGlyArgThrIleValIleAlaHisArgLeuSerThrVal 594
Db 1855 TTGAGAGGCGAGCTGAAGCGGAGCTTATCGCAAACTTGTGGAGGCTCAACGTATCAAT 2034
QY 595 ArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHis 614
Db 1915 AAAACGGCGCAACATTTGGTTCTCGTCAATGGCAAAATTCGTGAACGAATCTCAC 1974
QY 615 AspGluLeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGly 634
Db 1975 GATGAATGGTTGACCGGGAGGCGCTTATCGCAAACTTGTGGAGGCTCAACGTATCAAT 2034
QY 635 AsnGluIleGlu-----LeuGluAsnAlaThrGlyGluSerLysSerGluSerAsp 651
Db 2035 GAACAGAGGAAGTACGCGCTTGGAGGACCGCAGCTGAGGATCTCACGAATGCGAGAT 2094
QY 652 AlaLeuGluMetSerProLysAspSerGlySer-----Leu 664

Db 2095 ATTGCCAAATCAAAACTGCGTCAAGCGCATCATCCGATCTCGACGGAACCCACAACC 2154
QY 665 IleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlnAspArgLysLeu 684
Db 2155 ATTGACCGCAGCGGCCACCCACAAGTCTGTTCAGCGCGGATCTTCTTAAAGACCC--- 2211
QY 685 GlyThrLysGluAspLeuAsnGluAsnValProProValSerPheTrpArgIleLeuLys 704
Db 2212 -----CCGAAACAACCTCCGAAATACTCATTTAGGACGCTGCTCAAA 2253
QY 705 -----LeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 720
Db 2254 TTTGTTGCTTCTTCAACCGCCCTGAAATCCGTACATGCTCATCGTCTTGTCTCTCA 2313
QY 721 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 740
Db 2314 GTGTAGCTGGTGGTGGCCAAACCCAGCAGCATGTCTATATGCTAAAGCCATCAGACA 2373
QY 741 PheThrArgAspGluAspProGluThrLys---ArgGlnAsnSerAsnMetPheSerVal 759
Db 2374 CTCTCGCTCCAGAATCACAATATAGCAAGCTTCGACATGATCGGATTTCTGTCATTG 2433
QY 760 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 779
Db 2434 ATGTTCTTCTGTTGTTGATCATTTACGTTTATCAGCAGTCAACCAATGGTGTGCTATT 2493
QY 780 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 799
Db 2494 GCGTATGCTCCGAGAGACTTATTGTCGCGGAGAGACACTGCTTTCGAGGATACTC 2553
QY 800 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 819
Db 2554 CGTCAAGACATTTGTTCTTTTACAAAGAGAAATAGCACCGCGCTCTGACCTCTTTC 2613
QY 820 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 839
Db 2614 CTGTCACCGAGAGCAAGCATCTCCGGTGTAGCGGTGACTCTAGGCACGATCTG 2673
QY 840 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 859
Db 2674 ATGACCTCCAGCACCTTAGGAGCGCTATCATTTATGCCCTGGCGATTGGGTGGAATG 2733
QY 860 ThrLeuSerGlyGlnAlaLeuValProIleAlaIleAlaGlyValValGluMetLys 879
Db 2734 GCCTTAGTTGTATCTCGTGTGCGGTTCTCTCGGATCGCGTTCTACCGATTCTAT 2793
QY 880 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 899
Db 2794 ATGCTAGCCAGTTTCAATCACTCAGCTCCAAAGCTTGCTTATGAGGATCTGCAAACTTGTCT 2853
QY 900 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 919
Db 2854 TCGGAGGCTTACATCTCTATCCGACAGTTCGCTATTAAACCGGGAAGGATGCTGG 2913
QY 920 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 939
Db 2914 GAGATTACCTCCCGCTTACCCACAGCAGCAGCAGTCTAATCTCTCTCTTGAGG 2973
QY 940 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 959
Db 2974 TCATCCCTTTATATGCTGTCGCGACGACTTGTGTTTCTTCTGCGTCCGCTCGGGTTT 3033
QY 960 ArgPheGlyAlaTyrLeuValAlaAlaAsnGluPheMetAsnPheGlnAspValLeuVal 979
Db 3034 TGGTAGGAGGACACTTCTTGGTCACCGAGTATGACATTTTCCGCTTCTTCTTGT 3093
QY 980 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 999
Db 3094 TTCTCCGAGATCTCTTGTGCTCAATCCGCGGCGACGCTCTTCTTCTTCCACACAG 3153
QY 1000 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1019

Db	3154	ATGGCCAGGCGAAGAATCGCGCCGCCGAATTCGCGAGCACTGTCGACCGAAAGCCACAA	3213
Qy	1020	IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe	1039
Db	3214	ATTGATAACTGGTCTGAAGAGGCGAGAACTCGAAACGGTGTGAATCGAATTT	3273
Qy	1040	AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer	1059
Db	3274	AGGAACGTGCATTTACATACCCGACCCGCCAGAACAGCCTGCTCGCGCGCTTGCAC	3333
Qy	1060	LeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer	1079
Db	3334	CTGACCGTGAAGCGCTGCACAATATGTTCGCTGTTCGCAGCCAGCGGTGTGCGCAGAGT	3393
Qy	1080	ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp	1099
Db	3394	ACCACCAATGCATGCTTGAGCGCTTTACGATGCGATTCGTGCGGCTCCATCCTTGTGTGAT	3453
Qy	1100	GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer	1119
Db	3454	GGGAGGACATAAGTAACTAATATCAACTCCTACCCGACGCTTCTGTCACGTGGTCAGC	3513
Qy	1120	GlnGlnProIleLeuPheAspCysSerIleGluAlaAsnIleAlaTyrGlyAspAsnSer	1139
Db	3514	CAGGAGCGGACACTGTACCAGGCGCACCATCAAGGAACAATCTTACTTGGTATGTGCAG	3573
Qy	1140	ArgValValSerHisGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPhe	1159
Db	3574	GATGACGTACCGGAAGAATCTTGATTAAGGCTTGCAAGGACGCTAATATCTACGACTTC	3633
Qy	1160	IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer	1179
Db	3634	ATCATGTGCTCCCGAGGGCTTTAATACAGTTGTGGCAGCAAGGAGCGATGTGTCT	3693
Qy	1180	GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu	1199
Db	3694	GGCGGCCAAAGCAACGTGTGCCCATTCGCCGAGCCCTTCTCGGATCCCAAAATCCCT	3753
Qy	1200	LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla	1219
Db	3754	CTTCTCGATGAGCGAGTCACGCCCTCCGACTCCGAGCAGAAAGGTGTCACGCGCGCT	3813
Qy	1220	LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle	1239
Db	3814	TTGGATGCGCTGCCCGAGCGCGAACCACATTCGCCGTTCACACCGACTCACCCAGAT	3873
Qy	1240	GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis	1259
Db	3874	CAAAAGCGCGAGCTTATCTATGTTTTCACCAAGCAAGATCGTTCGAAAGCGGACGCCAC	3933
Qy	1260	GlnGlnLeuAlaGlnLysGlyIleTyrPheSerMetValSerValGlnAlaGlyAla	1279
Db	3934	AGCGAACTGGTCCAGAAAGGCGCGTACTACGAGCTGGTCAACTTGCAGAGCTGGGC	3993
Qy	1280	Lys 1280	
Db	3994	AAG 3996	

Search completed: November 6, 2002, 19:34:18
Job time : 383.556 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:33:09 ; Search time 20.063 seconds
(without alignments)
6135.192 Million cell updates/sec

Title: US-09-672-725C-27

Perfect score: 6473

Sequence: 1 MPEGRKGSAEKNEFKMGK.....LLAQKGIYFSNVQVQAGAKR 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PTR_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5866	90.6	1280	1 DVHU1	multidrug resistan
2	5653.5	87.3	1276	1 DVHYIC	multidrug resistan
3	5638	87.1	1276	2 A34786	multidrug resistan
4	5301	81.9	1276	1 DVMS1	multidrug resistan
5	5231.5	80.8	1277	2 JH0502	p-glycoprotein - r
6	4995	77.2	1104	1 DVMS1A	multidrug resistan
7	4922.5	76.0	1279	1 DVHU3	multidrug resistan
8	4870	75.2	1276	1 DVMS2	multidrug resistan
9	4829.5	74.6	1281	2 I48123	multidrug resistan
10	4816	74.4	1278	2 S41646	p-glycoprotein iso
11	4436	68.5	1287	2 S55692	p-glycoprotein - r
12	3221.5	49.8	1321	2 T42228	multidrug resistan
13	3217.5	49.7	1321	2 T42282	p-glycoprotein sis
14	2866	44.3	1294	2 T19982	bile salt transpor
15	2824.5	43.6	1275	2 T31073	hypothetical prote
16	2824	43.6	1289	2 D87789	multidrug resistan
17	2743	42.4	1321	2 T23476	protein C34G6.4 [I
18	2741	42.3	1321	2 S27337	multidrug resistan
19	2674.5	41.3	655	1 DVHY2C	multidrug resistan
20	2647	40.9	1283	2 A47377	multidrug resistan
21	2612.5	40.4	1286	2 T02187	probable ABC trans
22	2605.5	40.3	1292	2 T48007	P-glycoprotein hom
23	2601.5	40.2	1278	2 E86155	probable ABC trans
24	2568.5	39.7	1302	2 A41249	multidrug resistan
25	2532	39.1	1229	2 D85023	P-glycoprotein-lik
26	2529	39.1	1229	2 T52319	P-glycoprotein-lik
27	2527	39.0	1230	2 E85023	probable P-glycopr
28	2499	38.6	1302	2 B41249	multidrug resistan
29	2484	38.4	1229	2 F86155	probable ABC trans

30	2445	37.8	1408	2 T43261	multidrug resistan
31	2437.5	37.7	1323	2 H85202	hypothetical prote
32	2426	37.5	1310	2 S30328	multidrug resistan
33	2425	37.5	1286	2 A42150	p-glycoprotein ppp
34	2416	37.3	1266	2 T22090	hypothetical prote
35	2391.5	36.9	1288	2 T22094	hypothetical prote
36	2386	36.9	1302	2 S30327	multidrug resistan
37	2338.5	36.1	1254	2 S27338	p-glycoprotein C -
38	2330.5	36.0	1222	2 T14805	hypothetical prote
39	2306	35.6	1233	2 T04251	p-glycoprotein 2 -
40	2293	35.4	1254	2 T30855	multidrug resistan
41	2292	35.4	1245	2 G8404	probable P-glycopr
42	2290	35.4	1307	2 T30882	multidrug resistan
43	2276	35.2	1362	2 T41534	leptomycin B resis
44	2269.5	35.1	1318	2 T21266	hypothetical prote
45	2249	34.7	1247	2 F86405	probable P-glycopr

ALIGNMENTS

RESULT 1

DVHU1

multidrug resistance protein 1 - human

N;Alternate names: P-glycoprotein 1

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1990 #sequence_revision 18-Aug-1995 #text_change 19-Jan-2001

C;Accession: A34914; PS0162; S15500; A25059; S43838; I52238; I65204

R;Chen, C.; Clark, D.; Ueda, K.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.

J. Biol. Chem. 265, 506-514, 1990

A;Title: Genomic organization of the human multidrug resistance (MDR1) gene and origi

A;Reference number: A34914; MUID:90094448

A;Accession: A34914

A;Molecule type: DNA

A;Residues: 1-1280 <CHE>

A;Cross-references: GB:M29447; GB:J05168; NID:g187496; PIDN:AAA59576.1; PID:g386862

R;Kioka, N.; Yamano, Y.; Komano, T.; Ueda, K.

submitted to JIPID, April 1991

A;Reference number: PS0162

A;Accession: PS0162

A;Molecule type: DNA

A;Residues: 1-22 <KIO>

R;Kioka, N.; Yamano, Y.; Komano, T.; Ueda, K.

submitted to the EMBL Data Library, April 1991

A;Description: Transcriptional regulation of multidrug resistance gene (MDR1) express

A;Reference number: S15500

A;Accession: S15500

A;Molecule type: DNA

A;Residues: 1-22, 'R' <KI2>

A;Cross-references: EMBL:X58723; NID:g34522; PIDN:CAA41558.1; PID:g34523

R;Chen, C.; Chin, J.E.; Ueda, K.; Clark, D.P.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.

Cell 47, 381-389, 1986

A;Title: Internal duplication and homology with bacterial transport proteins in the m

A;Reference number: A25059; MUID:87028230

A;Accession: A25059

A;Molecule type: mRNA

A;Residues: 1-184, 'V', 186-1280 <CH2>

A;Cross-references: GB:M14758; NID:g187468; PIDN:AAA59575.1; PID:g307180

R;Chambers, T.C.; Pohl, J.; Glass, D.B.; Kuo, J.F.

Biochem. J. 299, 309-315, 1994

A;Title: Phosphorylation by protein kinase C and cyclic AMP-dependent protein kinase

A;Reference number: S43838

A;Accession: S43838

A;Molecule type: protein

R;Gekeler, V.; Weger, S.; Probst, H.

Biochem. Biophys. Res. Commun. 169, 796-802, 1990

A;Title: mdrl/P-glycoprotein gene segments analyzed from various human leukemic cell

A;Reference number: I52238; MUID:90290529

A;Accession: I52238

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 178-215 <RES>

A;Cross-references: GB:M37724; NID:g183537; PIDN:AAA88047.1; PID:g553314
A;Accession: I65204
A;Status: translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 800-856 <R2>
A;Cross-references: NID:g183538; PIDN:AAA88048.1; PID:g553315
C;Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic antitumor drugs.
C;Genetics:
A;Gene: GDB:PGV1; MDRI
A;Cross-references: GDB:120712; OMIM:171050
A;Map position: 7q21-7q21
C;Keywords: multidrug resistance protein; ATP-binding cassette homology
C;Superfamily: ATP; duplication; glycoprotein; nucleotide binding; P-loop; phosphoprotein;
F:1-638,653-1280/Region: duplication
F:49-350/Domain: hydrophobic <HB1>
F:351-637/Domain: hydrophobic <HL1>
F:410-604/Domain: ATP-binding cassette homology <ABC1>
F:427-434/Region: nucleotide-binding motif A (P-loop)
F:551-555/Region: nucleotide-binding motif B
F:638-708/Domain: linker <LIN>
F:709-993/Domain: hydrophobic <HB2>
F:994-1280/Domain: hydrophobic <HL2>
F:1053-1249/Domain: ATP-binding cassette homology <ABC2>
F:1070-1077/Region: nucleotide-binding motif A (P-loop)
F:1196-1200/Region: nucleotide-binding motif B
F:91,94,99/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:433/Binding site: ATP (Lys) #status predicted
F:661,667,671/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status exp
F:667,671,683/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status
F:1076/Binding site: ATP (Lys) #status predicted

Query Match 90.6%; Score 5866; DB 1; Length 1280;
Best Local Similarity 90.8%; Pred. No. 0;
Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;

Qy 1 MDPEGRKGS-A-EKNFWMKMGKSKKKEKPTVSTFAMFRYSNNWLDRLYMLVGTMAAI 59
Db 1 MDLEGRNGAKKKFKLNKS-EKDKKPKTVSVFMSFRYSNNWLDRLYMLVGTMAAI 59
Qy 60 IHGAALPLMLVGNMTDSTANAGISRNKTFPVINESITNNTQHFTHLEERTYAY 119
Db 60 IHGAALPLMLVGNMTDIFANAG-NLEDLSNITNRSNDINDTGFPMN-LEEDNTRYAY 117
Qy 120 YSGIGAGVLAAYIQVSWFCLAGROILKIRKQFFHAIMRQEGWFDVHDVDELNRLTD 179
Db 118 YSGIGAGVLAAYIQVSWFCLAGROILKIRKQFFHAIMRQEGWFDVHDVDELNRLTD 177
Qy 180 DVSKINEGIDKIGMFFQSIAITFTTGFVGTGKWLTVLILAIISPVLGSAAIWAKILS 239
Db 178 DVSKINEGIDKIGMFFQSIAITFTTGFVGTGKWLTVLILAIISPVLGSAAVWAKILS 237
Qy 240 SFTDKELLAYAKAGAAVEVLAARTVIAFGGKKELERYKNLEAKGIGIKKAITANI 299
Db 238 SFTDKELLAYAKAGAAVEVLAARTVIAFGGKKELERYKNLEAKRIGIKKAITANI 297
Qy 300 SIGAAFLIYASALAFWYGTSLVLSSEYITGQVLTIVFFSVLIGAFSIGQASPSIEAFN 359
Db 298 SIGAAFLIYASALAFWYGTSLVLSSEYISIGQVLTIVFFSVLIGAFSVGQASPSIEAFN 357
Qy 360 ARGAAEYFKIIDNKPISDTSYKSGHPDNITKGNLEFNHVFSPSRKVKILKGLNLKV 419
Db 358 ARGAAEYFKIIDNKPISDTSYKSGHPDNITKGNLEFNHVFSPSRKVKILKGLNLKV 417
Qy 420 QSGQVVALVNGSGGKSTTVQLMORLYDPTDGMVICDQDQDRTINVRHLREITGVVSOEP 479
Db 418 QSGQVVALVNGSGGKSTTVQLMORLYDPTDGMVSVSDQDQDRTINVRHLREITGVVSOEP 477
Qy 480 VLFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPKNKFTDLVGERGAQLSGGQK 539
Db 478 VLFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPKNKFTDLVGERGAQLSGGQK 537
Qy 540 RIARALVRNPKILLDEATSEALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADY 599

Db 538 RIARALVRNPKILLDEATSEALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADY 597
Qy 600 IAGFDGQVIVVEKGNHDELMEKEGIYFKLVTMTQTRGNEIELENATGESKESDALENSPKD 659
Db 598 IAGFDGQVIVVEKGNHDELMEKEGIYFKLVTMTQTRGNEIELENATGESKESDALENSND 657
Qy 660 SSGSLIKRRSTRRSIIHAPQODRKLTGKEDLNENVPVPSFWRLTKLNSWEPVYVVGIFC 719
Db 658 SRSLIKRRSTRRSVRSQADRKLTGKEDLNENVPVPSFWRLTKLNSWEPVYVVGIFC 717
Qy 720 AIINGLOPAFISFIISFIIGTRDEDPETKRONSMNMFSLVLVLGLIISITFFLAGQTF 779
Db 718 AIINGLOPAFISFIIGTRIDDPETKRONSNLFLSLFLALGLIISITFFLAGQTF 777
Qy 780 GKAGEILTCLRVMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSRSLAVIT 839
Db 778 GKAGEILTCLRVMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSRSLAVIT 837
Qy 840 QNIANLGTGIIISLIYQWQLTLLLLAIPIIATAGVEMKMLSGOALKDKKELEGAGKIA 899
Db 838 QNIANLGTGIIISFIYQWQLTLLLLAIPIIATAGVEMKMLSGOALKDKKELEGAGKIA 897
Qy 900 TEAENERTVVSLETRQKFEYVMAQSLQVPYRNSLRKAHIFGVFSFTQAMMYFSYAGCF 959
Db 898 TEAENERTVVSLETRQKFEYVMAQSLQVPYRNSLRKAHIFGVFSFTQAMMYFSYAGCF 957
Qy 960 RFGAYLVANEPNQDVLVLSAIVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSP 1019
Db 958 RFGAYLVAKLMSFEDVLLVFSAVVFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSP 1017
Qy 1020 IDSYSVPHGLKNTLEGNTVFNVEVFNPTRPDIPLVQLGLEVKKGQTLALVSGSGCGKS 1079
Db 1018 IDSYSVTEGMLNTEGNTVFNVEVFNPTRPDIPLVQLGLEVKKGQTLALVSGSGCGKS 1077
Qy 1080 TVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHLGLTVSQEPILFDCSIAENIAYGDN 1139
Db 1078 TVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHLGLTVSQEPILFDCSIAENIAYGDN 1137
Qy 1140 RVVSHIEIVQAAKEANIIHFETLPEKYNTRVGDKGTOLSGOKORTAIAARALVRQPHIL 1199
Db 1138 RVVSHIEIVQAAKEANIIHFETLPEKYNTRVGDKGTOLSGOKORTAIAARALVRQPHIL 1197
Qy 1200 LDEATSEALDTESEKVVQEQALDKAREGRTICIVIAHRLSTIONADLIIVFONGKVEHGT 1259
Db 1198 LDEATSEALDTESEKVVQEQALDKAREGRTICIVIAHRLSTIONADLIIVFONGKVEHGT 1257
Qy 1260 QQLLAQKGIYFSMVSVQAGAKR 1281
Db 1258 QQLLAQKGIYFSMVSVQAGTKR 1279

RESULT 2
DVH1C
Multidrug resistance protein 1 - Chinese hamster
N:Alternate names: p-glycoprotein pgp1
C:Species: Crictetus glisus (Chinese hamster)
C:Date: 31-Dec-1990 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A38696; B38696; A27126; S33768; I52823
R:Devine, S.B.; Hussain, A.; Davide, J.P.; Melera, P.W.
J. Biol. Chem. 266, 4545-4555, 1991
A:Title: Full length and alternatively spliced pgp-1 transcripts in multidrug-resista
A:Reference number: A38696; MUID:91154265
A:Accession: A38696
A:Molecule type: mRNA
A:Residues: 1-1276 <DE>
A:Cross-references: GB:M59253; NID:g191154; PIDN:AAA37004.1; PID:g191155
A:Accession: C38696
A:Molecule type: mRNA
A:Residues: 108-1276 <DE1>
A:Cross-references: GB:M59254; NID:g191156; PIDN:AAA37005.1; PID:g191157
A:Experimental source: clone ADX185
A:Accession: B38696

A:Title: Cell-specific activity of cis-acting regulatory elements in the promoter of the
A:Reference number: I57510; MUID:91042535
A:Accession: I57510
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-21 <RES>
A:Cross-references: GB:M60348; NID:g199102; PIDN:AAA39513.1; PID:g554199
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: mdr1 (pgpl)
A:Introns: 21/2; 38/3; 92/1; 111/2; 175/2; 233/3; 274/2; 332/3; 370/3; 407/3; 449/3; 517
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane pr
F:1-637,653-1276/Region: duplication
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-633/Region: nucleotide-binding motif A (P-loop)
F:550-554/Region: nucleotide-binding motif B
F:1051-1247/Domain: ATP-binding cassette homology <ABC2>
F:1068-1075/Region: nucleotide-binding motif A (P-loop)
F:1194-1198/Region: nucleotide-binding motif B
F:713,91,96,103/Binding site: carboxylate (Asn) (covalent) #status predicted
F:432/Binding site: ATP (Lys) #status predicted
F:1074/Binding site: ATP (Lys) #status predicted

Query Match 81.9%; Score 5301; DB 1; Length 1276;
Best Local Similarity 80.3%; Pred. No. 2.4e-298;
Matches 1030; Conservative 130; Mismatches 114; Indels 8; Gaps 5;

QY	1	MDPEGGKSAENFWMKSKKKKKKKKPTVTFAMFRYSNMLDRLMLVGTMAAII	60
DB	1	MEFEENLGRADNFKSMKKKS-KKEKKEKPAVGFGFMYADWLDKLCMLGLTAAII	59
QY	61	HGAALPLMLVFGNMTDSANAGISRNKTFPPVLINESINNTQHIN-HLEEWYIAY	119
DB	60	HGTLPLMLVFGNMTDSFKAEAS---ILPSITNQGSPNSTLIISNSLEEEMAIYAY	116
QY	120	YSGIGAGVLVAIVQSVFCLAGROILKIRKOFFHAIMROBIFGDFVDVDELNRLTD	179
DB	117	YTGIGAGVLIVAYIQSVLWCLAGROILKIRKOFFHAIMRQBFGDFVDVDELNRLTD	176
QY	180	DVSKINEGIDKIGMEFFOSTAFFTGIVGFTRGWKLILVILAISPVGLSAAIWAKILS	239
DB	177	DVSKINDGIDKIGMEFFQSITFLAGFIIGFTSGWKLILVILASPLIGLSALWAKYLT	236
QY	240	SFTDKELLAYAKAGAAEVLAAIRTVIAFGQKKELERYKNLEAKGIGIKKAITANI	299
DB	237	SFTNKLQAYAKAGAAEVLAAIRTVIAFGQQQKELERYKNLEAKNKGIKKAITASI	296
QY	300	STGAPELLIYASALAFWYGTSLVLSSEYTIQGVLTFFSVLIGAFSGQASPSIEAFAN	359
DB	297	STGIAYLLVYASALAFWYGTSLVLSNEYSIGEVLTVFSSILGTFSGHLPANIEAFAN	356
QY	360	ARGAAVEIFKIIDNKPSIDYSKSGHKPDNKGNLKGNLFNHFSPSRKEVKILKGLNKV	419
DB	357	ARGAAVEIFKIIDNPKPSIDFSFKYKPDNKGNLKGNLFNHFNPNPSRSEVOILKGLNKV	416
QY	420	QSGQTVALVNGSGCKSTTVQLMORLYDPTDGMVCDIGDQDRTINVRHLREITGVVQEP	479
DB	417	KSGQTVALVNGSGCKSTTVQLMORLYDPLGVSVISDQDRTINVRVLRRIIGVVQEP	476
QY	480	VLFATTIANIRYGRNVTMBIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKQ	539
DB	477	VLFATTIAENIRYGRNVTMBIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKQ	536
QY	540	RTAIAALVRNPKILLDEATSEALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADY	599
DB	537	RTAIAALVRNPKILLDEATSEALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADY	596
QY	600	IAGFDGVTVEKGNHDELMKKEGIYFKLVMTQTRGNELELNATGESKSDALEMSPKD	659
DB	597	IAGFDGVTVEKGNHDELMKKEGIYFKLVMTQTRGNELELNATGESKSDALEMSPKD	656

QY	660	SGSSLIKRRSTRRSIHAPOQODRKLGTKEDLNENVPVPSFWRLIKLNSTEWPFYVVGIFC	719
DB	657	SKSPLI-RRSIYRSVHRKQDQERRLSMKEAVDEVDPLVSEFWRLINLNLSEWPLYLVGLC	715
QY	720	AIINGGLQAFSIIIFRIIGFTREDPEKRONSNWFSVLRLVLGLIISFIIFLQGTTF	779
DB	716	AVINGCQIPVPAIVFSRIVGVFSDDDHETKRONCLNLSLFFLVMLGLISFVIYFFQGTTF	775
QY	780	GKAGEILTKLRVMVFRSMRLQDVSWFDDPKNTTGALTTRLANDAAQVGAIGSLAVIT	839
DB	776	GKAGEILTKRVYVMVFKSMLRQDISWFDHDKNSTGSLTTRLASDASSVKGAMGARLVV	835
QY	840	QNIANLGTGIIISLITGWLITLILLAIPIIATAGVVEKMLSGQALKDKKLEAGKTA	899
DB	836	QNVANLGTGILSVYGWQUTLLVWIPIVLGGIEMKLLSGQALKDKKLEISGKTA	895
QY	900	TEAIENFRTVSLTREOKFEYMAQSLQVPRNSLRKAHFGVFSFISITAMVFSYAGCF	959
DB	896	TEAIENFRTIVSLTREOKFETMTAQSLQVPRNMAKKAHVGITFTQAMVFSYACF	955
QY	960	RFGAYLVANEFMFQDVLVFSIAIVFGAMAVQVSSFADYAKAKVSAHVIMIIEKSP	1019
DB	956	RFGAYLVAAQQLMTFENVMVFSVAVFEAGMAAGNTSSFADYAKAKVSAHIIIEKTP	1015
QY	1020	IDSVSPHGLKPNTEGNTFNEVFNYPTRDPVLOGLSLEVKKGQILALVSSGCGKS	1079
DB	1016	IDSVTEGLKPTLLEGNVAFNGVQFNYPTRPNIPVLOGLSLEVKKGQILALVSSGCGKS	1075
QY	1080	TVVQLLERFDPLAGSVLDGKRIKHLNVQWLRHAHLGIVSQBPILFDCSIAENIAYGNS	1139
DB	1076	TVVQLLERFDPMAGSVLDGKRIKHLNVQWLRHAHLGIVSQBPILFDCSIAENIAYGNS	1135
QY	1140	RVYSHEEIVQAAKEANIHFIETLPEKYNTRVGDGTQSLSGQKQRIARALVRQPHIL	1199
DB	1136	RAYSHEEIVRAAKEANIHQFIDSLDPKYNTRVGDGTQSLSGQKQRIARALVRQPHIL	1195
QY	1200	LLDEATSAIDTESEKVVQALDKAREGRTCIIVIAHRLSTIONADLIVVFQNGKVEHGT	1259
DB	1196	LLDEATSAIDTESEKVVQALDKAREGRTCIIVIAHRLSTIONADLIVVIENGKVEHGT	1255
QY	1260	QQLLAQKGIYFMSVQAGAKR 1281	
DB	1256	QQLLAQKGIYFSM--VQAGAKR 1275	

RESULT 5
JH0502

p-glycoprotein - rat
N:Alternate names: multidrug resistance protein mdr1b
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 02-Feb-2001
C:Accession: JH0502; S22353; S22352
R:Silverman, J.A.; Raulio, H.; Gant, T.W.; Thorgerirsson, S.S.
Gene 106, 229-236, 1991
A:Title: Cloning and characterization of a member of the rat multidrug resistance (mdr1) gene
A:Reference number: JH0502; MUID:92039081
A:Accession: JH0502
A:Molecule type: mRNA
A:Residues: 1-1277 <STL>
A:Cross-references: GB:M62425
R:Deuchars, K.L.; Duthie, M.; Ling, V.
Biochim. Biophys. Acta 1130, 157-165, 1992
A:Title: Identification of distinct P-glycoprotein gene sequences in rat.
A:Reference number: S22351; MUID:92223089
A:Accession: S22353
A:Molecule type: DNA
A:Residues: 1212-1226, 'I', 1228-1277 <DED>
A:Cross-references: EMBL:X61104; NID:g56890; PIDN:CAA43416.1; PID:g1334219
A:Accession: S22352
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1212-1226, 'I', 1228-1270, 'SV', 1271-1277 <DB2>
A:Cross-references: EMBL:X61103; NID:g56888; PIDN:CAA43415.1; PID:g1334218

C:Comment: This protein has the nucleotide binding motifs and ATP binding active transpo

C:Genetics:

A:Gene: mdr1b

C:Superfamily: multidrug resistance protein; ATP-binding cassette homology

C:Keywords: ATP; duplication; glycoprotein; membrane protein; nucleotide binding; P-loop

F:409-603/Domain: ATP-binding cassette homology <ABC1>

F:426-434/Region: nucleotide-binding motif A (P-loop)

F:550-554/Region: nucleotide-binding motif B

F:1053-1248/Domain: ATP-binding cassette homology <ABC2>

F:1195-1199/Region: nucleotide-binding motif A (P-loop)

F:1070-1078/Region: nucleotide-binding motif B

F:432/Binding site: ATP (Lys) #status predicted

F:1076/Binding site: ATP (Lys) #status predicted

Query Match 80.8%; Score 5231.5; DB 2; Length 1277;
Best Local Similarity 80.0%; Pred. No. 2.5e-294;
Matches 1029; Conservative 131; Mismatches 111; Indels 15; Gaps 9;

```
QY 1 MDPEGRKSAEKNFWMGKKKKKKKKKPTVSTFAMFRYSNWLDRMLVGTMAAII 60
DB 1 MEEFELGRADKNFSGMKGKKKK--EKEKKPAVGIFGFRYADWLDKLCMALGTLAAII 58
QY 61 HGAALPLMLVFGNMTDSPANAGISRN-KTFPVIINESTNNTQHPIN--HLEEMTYIA 117
DB 59 HGTLLPLMLVFGYMTDSFTP---SRDPHSDRAITNQSEINST-HTVSDTSLIEDMAMYA 114
QY 118 YYTSGTGAGVLVAAVYQVSEFWCLAAAGROILKIRKQFFHAIMQOEIGWFDVHVGELNTRL 177
DB 115 YYTGTGAGVLVYAYVQVSLWCLAAAGROIHKIRKQFFHAIMQOEIGWFDVNDAGELNTRL 174
QY 178 TDDVSKINEGIGDKIGMFFQSTATFTFTGTVGTRGKWLTLVLAISPVLGLSAAIWKAI 237
DB 175 TDDVSKINDIGDKLGMFQSTFTFAGFGGKKELERYNKNLEAKRGVGIKKAITA 234
QY 238 LSSTFKELLAYAKAGAAVEVLAAITRTVIAFGGKKELERYNKNLEAKRGVGIKKAITA 297
DB 235 LTFSTNKLQAYAKAGAAVEVLAAITRTVIAFGGKKELERYNKNLEAKRGVGIKKAITA 294
QY 298 NISGAFLIYASYALAFWYGSVLISSEYITIGQVLTFFVFSLIGAFSIGQASPSIEAF 357
DB 295 NISGIIAYLVYASYALAFWYGSVLISSEYISIGQVLTFFVFSILLOTFSIGHLAPNIEAF 354
QY 358 ANARGAAEYFIKIIDNKPISDYSKSGHKPDNTKGNLFKNVHFSPSKVEKILKGLNL 417
DB 355 ANARGAAEYFIKIIDNEPSIDFSTGKHKPDSIMGNLFKNVHFSPSKVEKILKGLNL 414
QY 418 KVQSGQTVLVGNSGCGKSTTVOLMORLYDPTDGMVCIDGQDITINVRHLREITGVWSQ 477
DB 415 KVQSGQTVLVGNSGCGKSTTVQLLQRLYDPIEGEWSIDGQDITINVRHLREITGVWSQ 474
QY 478 EPLVFATTIENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQ 537
DB 475 EPLVFATTIENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQ 534
QY 538 KQRTAARALVRNPKILLDEATSALDTESEAVVQVALDKARGRTIIVIAHRLSVRNA 597
DB 535 KQRTAARALVRNPKILLDEATSALDTESEAVVQVALDKARGRTIIVIAHRLSVRNA 594
QY 598 DVIAGFDGVTVEKGNHDELMKEGIIYFKLVMTQTRGNEILENATGESKESDALEMS 657
DB 595 DVIAGFDGVTVEKGNHDELMKEGIIYFKLVMTQTRGNEILENATGESKESDALEMS 654
QY 658 KDSGSLTKRRSTRSIIHAPQGDRIKLTGKEDLNENPVPSFVRILKLNTEPFFVVG 717
DB 655 EESKSPIL-RRSIRRSIHRRODQERRLSKEDVDVDPVMSFWQILKLNISEWPLYVVG 713
QY 718 FCAINGGLQAPAFSIISRIIGITREDDEPTKQNSNMESVLFLVGLIISFTFFLQ 777
DB 714 LCAVINGCIQVFAIVFSKIVGVFSRDOHETKORNCNLNLSFLVGMGLISFVYFFQ 773
QY 778 TFGKAGETILTKRLRYMVRSMRQDVSWFDPPKNTTGTALTRLANDAAQVKGAGISRLAY 837
DB 774 TFGKAGETILTKRLRYMVRSMRQDVSWFDPPKNTTGTALTRLANDAAQVKGAGISRLAY 833
```

```
QY 838 ITONIANLGRGIIS--LIYGWQLTLLLAIVPIIATAGVYVEMKMLSGQALKDKKELEGA 895
DB 834 VTQNVANLGTGIIILSLVLVYGWQLTLLLVIIPLVLGGIEMKLLSGQALKDKKELEIS 893
QY 896 GKIAATEAENFRVTUSTRQKFEYMYAQSLQVYRNSLRKAHIFGVSVSITQAMVFSY 955
DB 894 GKIAATEAENFRVTUSTRQKFEYMYAQSLQVYRNSLRKAHIFGVSVSITQAMVFSY 953
QY 956 AGCFRGAYLVANEFMFQDVLVFSVAVFGAMAVGVSSVAFADYAKAKVSAAHVIMIE 1015
DB 954 AACFRGAYLVARELMTFENVMVLSAVVFGAMAGNTSSVAFADYAKAKVSAASHIIGIE 1013
QY 1016 KSLPLDISYSPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKQOTLALVGSSG 1075
DB 1014 KIPEIDSYSTEGKPNWLEGNVKNFNGVKNFYPTRPNIPVLQGLSFEVKKQOTLRLVGSSG 1073
QY 1076 CGKSTVVQLLERFDYPLAGSVLIDGKEIKHLNVQWLAHLGIYVSOEPILEDGCSAENIAY 1135
DB 1074 CGKSTVVQLLERFDYPLAGSVLIDGKEIKHLNVQWLAHLGIYVSOEPILEDGCSAENIAY 1132
QY 1136 GDSRVVSHEEIVQAQAEANIHFIETLPEKYNTRVGDGTQSLGGQKQRIATARALVRQ 1195
DB 1133 GDSRVVSHEEIVRAAREANIHQFIDSLPEKYNTRVGDGTQSLGGQKQRIATARALVRQ 1192
QY 1196 PHILLDEATSALDTESEKVVQEALDKARGRTCIVIAHRLSTIONADLIIVFQNGKVKE 1255
DB 1193 PHILLDEATSALDTESEKVVQEALDKARGRTCIVIAHRLSTIONADLIIVFQNGKVKE 1252
QY 1256 HGTHQOLLAOKGIYFSVMVSVQAGAKR 1281
DB 1253 HGTHQOLLAOKGIYFSM--VQAGAKR 1276
```

RESULT 6
DMSIA
multidrug resistance protein la - mouse (fragment)
N:Alternate names: P-glycoprotein la
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: A34175
J. Hsu, S.I.H.; Lothstein, L.; Horwitz, S.B.
A:Title: Differential overexpression of three mdr gene family members in multidrug-re
A:Reference number: A34175; MUID:89308614
A:Accession: A34175
A:Molecule type: mRNA
A:Residues: 1-1104 <HSU>
A:Cross-references: GB:J04839; NID:g200329; PIDN:AAA03243.1; PID:g200330
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c
structurally and functionally unrelated lipophilic anticancer drugs.
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane
F:234-428/Domain: ATP-binding cassette homology <ABC1>
F:251-258/Region: nucleotide-binding motif A (P-loop)
F:375-379/Region: nucleotide-binding motif B
F:877-1073/Domain: ATP-binding cassette homology <ABC2>
F:894-901/Region: nucleotide-binding motif A (P-loop)
F:1020-1024/Region: nucleotide-binding motif B
F:1070-1078/Region: nucleotide-binding motif A (P-loop)
F:257/Binding site: ATP (Lys) #status predicted
F:900/Binding site: ATP (Lys) #status predicted

Query Match 77.2%; Score 4995; DB 1; Length 1104;
Best Local Similarity 89.3%; Pred. No. 9.5e-281;
Matches 985; Conservative 61; Mismatches 57; Indels 0; Gaps 0;

```
QY 179 DVVSKINEGIGDKIGMFFQSTATFTFTGTVGTRGKWLTLVLAISPVLGLSAAIWKAIL 238
DB 1 DVVSKINEGIGDKIGMFFQSTATFTFTGTVGTRGKWLTLVLAISPVLGLSAAIWKAIL 238
QY 239 SFTDKELLAYAKAGAAVEVLAAITRTVIAFGGKKELERYNKNLEAKRGVGIKKAITA 298
DB 61 SFTDKELLAYAKAGAAVEVLAAITRTVIAFGGKKELERYNKNLEAKRGVGIKKAITA 298
```

QY	299	ISIGAEFLIYASYALAFWYGTSLVLSSEYTIQOVLTVPFSVLIGAFSIGQASPSIEAFA	358
Db	121	ISMGAPELLIYASYALAFWYGTSLVLSKESIQOVLTVPFSVLIGAFSIGQASPSIEAFA	180
QY	359	NARGAAEIKIDNKPSIDSYSKSGHKPDNIKGNLEFNHFSYPSRKEVILKGLNLK	418
Db	181	NARGAAEYEVFKIDNKPSIDSFSGHKPDNIQGNLEFNHFSYPSRKEVQILKGLNLK	240
QY	419	VOSGOTVALVNGSCGKSTTVQLMORLYDPTDGMVCIQDQDITINVRHLREITGVWSQE	478
Db	241	VKSQOTVALVNGSCGKSTTVQLMORLYDPLDGMVSIQDQDITINVRHLREIIGVWSQE	300
QY	479	PVLFAITIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPKQFDTLVGERGAOLSGGOK	538
Db	301	PVLFAITIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPKQFDTLVGERGAHVSOGGOK	360
QY	539	ORTAIARALVRNPKILLDEATSAIDTESAEVVQVALDKARKGRTTIVIAHRLSTVRNAD	598
Db	361	ORTAIARALVRNPKILLDEATSAIDTESAEVVQVALDKARKGRTTIVIAHRLSTVRNAD	420
QY	599	VIAGFDGVIYVEKGNHDELMKEGIYFKLVMTQTRGNETELENAATGESKESDALEMSPK	658
Db	421	VIAGFDGVIYVEKGNHDELMKEGIYFKLVMTQTAGNETELENAACKSKDEIDNLDMSK	480
QY	659	DGSSSLIKRSTRRSIHAPQGDRLKLTGKEDLNENVPVSWRILKLNSTWPFVVGIF	718
Db	481	DGSSSLIRSTRSRKSCGPHDQDRKLTKEALDEDVPPASFWRILKLNSTWPFVVGIF	540
QY	719	CAIINGLOPAPSIIEFRIGITFTRDEDPETQRONMPSVLFLVLGIISFTIFFLQGT	778
Db	541	CAIINGLOPAPSVIFKSVFVGTNGPPETQRQNSLFLSLFLILGIISFTIFFLQGT	600
QY	779	FGKAGEILKRLRYMFRSMLRQDYSWFDDPKNTTGALTTRLANDAAOVKAIGSLAVI	838
Db	601	FGKAGEILKRLRYMFKSMLRQDYSWFDDPKNTTGALTTRLANDAAOVKATGSRUAVI	660
QY	839	TONIANLGTGIIISLIYGNQLFLLLAIVPIIAIAGVYEMKLSQALKRKELEGAGKI	898
Db	661	QONIANLGTGIIISLIYGNQLTLLLAIVPIIAIAGVYEMKLSQALKRKELEGSGKI	720
QY	899	ATEAIENFTVSLTRREQEFYMYAQSLQVYPRNSLRKAHFGVSFTQAMTFSYAGC	958
Db	721	ATEAIENFTVSLTRREQEFYMYAQSLQIPIYRNAMKKAHVFGITFSFTQAMTFSYAA	780
QY	959	FRFGAYLVANEPFNFQDVLVFSALVFGAMAVGVSSFAPDYAKAKYSAAHVIMIEKSP	1018
Db	781	FRFGAYLVYQQLMTFENVLLVFSALVFGAMAVGVSSFAPDYAKATVSASHIIRIIEKTP	840
QY	1019	LIDSYSPLGKPKNTLEGWTFNEVFNYPTRPIDPVLOGLSLEVKGQTALVSGSCGK	1078
Db	841	EIDSYSTQGLKPNMLEGNVQFSGVFNYPTRPSIPVLOGLSLEVKGQTALVSGSCGK	900
QY	1079	STWVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHAHLGIVSQEPIILDCSTAENIAYGDN	1138
Db	901	STWVOLLERFYDPMAGSVFLDGKEIKQLNVQWLRHAHLGIVSQEPIILDCSTAENIAYGDN	960
QY	1139	SRVVSHEEIVQAAKENAIHHFIETLPEKYNTRVGDKGTQLSGGQKRIATARALVRPHI	1198
Db	961	SRVVSHEEIVRAAKEANIHQFTDSLDPKYNTRVGDKGTQLSGGQKRIATARALVRPHI	1020
QY	1199	LILDEATSAIDTESKVVQEALDKAREGRTTIVIAHRLSTIONADLIWVFQNGKVBHGT	1258
Db	1021	LILDEATSAIDTESKVVQEALDKAREGRTTIVIAHRLSTIONADLIWVFQNGKVBHGT	1080
QY	1259	HQOLLAQKGIYFSMVSVOAGAKR	1281
Db	1081	HQOLLAQKGIYFSMVSVOAGAKR	1103
RESULT 7			
DYHU3			
multidrug resistance protein 3 - human			

N:Alternate names: P-glycoprotein MDR3			
C:Species: Homo sapiens (man)			
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #Text_change 19-Jan-2001			
C:Accession: J50051; S01346; A42213; I38015			
R:van der Blik, A.M.; Koolman, P.M.; Schneider, C.; Borst, P.			
Gene 71, 401-411, 1988			
A>Title: Sequence of mdr3 cDNA encoding a human P-glycoprotein.			
A:Reference number: J50051; MUID:89138016			
A:Accession: J50051			
A:Molecule type: mRNA			
A:Residues: 1-1279 <VA1>			
R:van der Blik, A.M.; Baas, F.; ten Houte de Lange, T.; Koolman, P.M.; van der Velde			
EMBO J. 6, 3325-3331, 1987			
A>Title: The human mdr3 gene encodes a novel P-glycoprotein homologue and gives rise			
A:Reference number: S01346; MUID:8811519			
A:Accession: S01346			
A:Molecule type: mRNA			
A:Residues: 856-1093, 'FVDFGFO', 1094-1279 <VA2>			
A:Cross-references: EMBL:X06181; NID:934524; PIDN:CAA29547.1; PID:934525			
R:Lincker, C.R.; Smic, J.J.M.; van der Velde-Koerts, T.; Borst, P.			
J. Biol. Chem. 266, 5303-5310, 1991			
A>Title: Structure of the human MDR3 gene and physical mapping of the human MDR locus			
A:Reference number: A42213; MUID:91161629			
A:Accession: A42213			
A>Status: not compared with conceptual translation			
A:Molecule type: DNA			
A:Residues: 1-1279 <LN1>			
R:Smic, J.J.; Mol, C.A.; van Deemter, L.; Wagenaar, E.; Schinkel, A.H.; Borst, P.			
Biochim. Biophys. Acta 1261, 44-56, 1995			
A>Title: Characterization of the promoter region of the human MDR3 P-glycoprotein gen			
A:Reference number: I38015; MUID:95200972			
A:Accession: I38015			
A>Status: translated from GB/EMBL/DBJ			
A:Molecule type: mRNA			
A:Residues: 1-61, 'RGSSRVDIQAC' <RES>			
A:Cross-references: EMBL:Z35284; NID:91005662; PIDN:CAA84542.1; PID:91006663			
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c			
tructurally and functionally unrelated lipophilic antitumor drugs.			
C:Genetics:			
A:Gene: GDB:PGY3; MDR3			
A:Cross-references: GDB:120713; OMIM:171060			
A:Map position: 7q21-7q21			
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology			
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane			
F:1-640,653-1279/Region: duplication			
F:412-606/Domain: ATP-binding cassette homology <ABC1>			
F:429-436/Region: nucleotide-binding motif A (P-loop)			
F:553-557/Region: nucleotide-binding motif B			
F:638-694/Domain: linker <LINK>			
F:1053-1248/Domain: ATP-binding cassette homology <ABC2>			
F:1069-1076/Region: nucleotide-binding motif A (P-loop)			
F:1193-1199/Region: nucleotide-binding motif B			
F:91.97/Binding site: carbohydrate (Asn) (covalent) #status predicted			
F:435/Binding site: ATP (Lys) #status predicted			
F:1075/Binding site: ATP (Lys) #status predicted			
Query Match			
Best Local Similarity 76.08; Score 4922.5; DB 1; Length 1279;			
Matches 977; Conservative 129; Mismatches 163; Indels 19; Gaps 9;			
QY	1	MDPEGGRKG-----SAEKNFWMKSKKKEKKK-PTVSTFAMFPRYSNWLDRLYMLV	53
Db	1	MDLEAKNGTAWRPTSAEGDF-ELGISKKQKTKTKVKMIGVLTLPYSDWQDKLFMSL	59
QY	54	GTMAAIIHGAALPLMLVFGNMTDSFANAGISRNKTPFVINESITNTQHFHINLHEEM	113
Db	60	GTMAIAHSGSLPLMMIVFGEMTKFVDT--AGNFSFPVNFSLSLNPGK----ILEEM	113
QY	114	TTYAYYSSIGAGVLVAAYIQVSEFWCLAAAGROILKIRKQFFHAIMRQETGFWHDVGEL	173
Db	114	TRYAYYSGLGAGVLVAAYIQVSEFWTLLAAGROIRKIRKQFFHAILRQETGFWFDINDTEL	173

174 NTRLRDDVSKINNEGIDKIGMFFQSIATFFTGFIQVFTGRGKLTILVILAIISPVGLGSAAI 233
Db 174 NTRLRDDISKISIEGIDKGVGMFFQAVFFAGFIQVFTGRGKLTILVILAIISPVGLGSAAV 233
QY 234 WAKILSSFTDKELLAYAKAGAAVEELAAIRTVIAFGQKKELERYKNKNLEAEKIGIKK 293
Db 234 WAKILSFAFDKELAAAYAKAGAAVEELAAIRTVIAFGQKKELERYQKHLENAKEIGIKK 293
QY 294 AITANISIGAAFLLIYASVALAFWYGTSLVLSSEYITIGQVLTVFSSVLIGAFSGQASPS 353
Db 294 AITANISMGIAFLLIYASVALAFWYGTSLVLSSEYITIGQVLTVFSSVLIGAFSGQAAAPC 353
QY 354 IAFANARGAAVEIFKIDNKSIDSYSKSGHKPDNIKGNLEFKNVHESYPSRKEVKILK 413
Db 354 IAFANARGAAVEIFDIIDNKNIDKIDSFSGHGPDSIKGNLEFNDVHESYPSRANVKILK 413
QY 414 GLNLKVQSGQVALVNSCGSTTVQLMORLYDPTDGMVCIDGQDITINVRHLREITG 473
Db 414 GLNLKVQSGQVALVNSCGSTTVQLMORLYDPTDGMVCIDGQDITINVRHLREITG 473
QY 474 VVSQEPVLFTATTAEINIRYGRNVMTDEIEKAVKANAYDFIMKLPKFDILVGERQAL 533
Db 474 VVSQEPVLFTATTAEINIRYGRNVMTDEIEKAVKANAYDFIMKLPKFDILVGERQAL 533
QY 534 SGGOKORIAALVNPKNILLDEATSAIDTSEAVVOVALDKARKGRTTIVIAHRLST 593
Db 534 SGGOKORIAALVNPKNILLDEATSAIDTSEAVVOVALDKARKGRTTIVIAHRLST 593
QY 594 VRNADVIAGFDGVIIVKGNHDELMEKGIYFKLVMTQRMNEIELENAITGESKESDAL 653
Db 594 VRNADVIAGFDGVIIVKGNHDELMEKGIYFKLVMTQRMNEIELENAITGESKESDAL 653
QY 654 EMPKSGSSSLIKRSTRSIHAPOQDRKLTGKED-LNENVPVPSFWILKLNSTWEPY 712
Db 654 EMPKSGSSSLIKRSTRSIHAPOQDRKLTGKED-LNENVPVPSFWILKLNSTWEPY 712
QY 713 FVYGIFCAINGLQAPAFSIFRIIGTIRDEDPETKRNQSMFSLVLVLGIISFIFF 772
Db 713 FVYGIFCAINGLQAPAFSIFRIIGTIRDEDPETKRNQSMFSLVLVLGIISFIFF 772
QY 773 FLOGFTFGKAGEILTKRLRVWFRSMRLQDVSFDFDPKNTTGALTRLANDAAQYKATG 832
Db 773 FLOGFTFGKAGEILTKRLRVWFRSMRLQDVSFDFDPKNTTGALTRLANDAAQYKATG 832
QY 833 SRLAVITONIANLGTGIIISLYGWOLTLALLAIPIAIVAGVEMKLSGQALKDKKEL 892
Db 833 SRLAVITONIANLGTGIIISLYGWOLTLALLAIPIAIVAGVEMKLSGQALKDKKEL 892
QY 893 EGAKIATEAENFRVTVSLTREKFEYMYAQSLQVPRYNSLRKAHIFGVSPSITOAMY 952
Db 893 EGAKIATEAENFRVTVSLTREKFEYMYAQSLQVPRYNSLRKAHIFGVSPSITOAMY 952
QY 953 FSYAGCFRFGAYLVNFMNFQDVLVFSIAIVFGMAVGOVSSFADYAKAKYSAAHVIM 1012
Db 953 FSYAGCFRFGAYLVNFMNFQDVLVFSIAIVFGMAVGOVSSFADYAKAKYSAAHVIM 1012
QY 1013 IIEKSPIDISYSPHGLKPNLTLEGNTVEVFNYPTRPDIPVLQGLSLEVKKGQTLALVG 1072
Db 1013 IIEKSPIDISYSPHGLKPNLTLEGNTVEVFNYPTRPDIPVLQGLSLEVKKGQTLALVG 1072
QY 1073 SSGCGKSTVVOLLBRFVDPYLAGSVLIDGKEIKHLNVOWLRAHGIYVQSEPIEDCSIAEN 1132
Db 1073 SSGCGKSTVVOLLBRFVDPYLAGSVLIDGKEIKHLNVOWLRAHGIYVQSEPIEDCSIAEN 1132
QY 1133 IAYGNSRVVSHETVQAAEANTHETLPEKYNTRVGDGKGTQSLSGGQKQRIATARAL 1192
Db 1133 IAYGNSRVVSHETVQAAEANTHETLPEKYNTRVGDGKGTQSLSGGQKQRIATARAL 1192
QY 1193 VRQPHILLDDATSAIDTSEKVVQEAALDKAREGRTCVIAHRLSTIONADLIIVFONGK 1252
Db 1193 VRQPHILLDDATSAIDTSEKVVQEAALDKAREGRTCVIAHRLSTIONADLIIVFONGK 1252
QY 1253 VKEHGTQOOLLAQKGIYFSMVSVQAGTQ 1277
Db 1253 VKEHGTQOOLLAQKGIYFSMVSVQAGTQ 1277

1250 VKEHGTQOOLLAQKGIYFSMVSVQAGTQ 1277
Db 1250 VKEHGTQOOLLAQKGIYFSMVSVQAGTQ 1277
RESULT 8
DMS2
multidrug resistance protein 2 - mouse
N:Alternate names: p-glycoprotein MDR2
C:Species: Mus musculus (house mouse)
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: A30409; S70711
R:Gros, P.; Raymond, M.; Bell, J.; Housman, D.
Mol. Cell. Biol. 8, 2770-2778, 1988
A:Title: Cloning and characterization of a second member of the mouse mdr gene family
A:Reference number: A30409; MUID:88302195
A:Accession: A30409
A:Molecule type: mRNA
A:Residues: 1-1276 <HSU>
A:Cross-references: GB:J03398; NID:g199109; PIDN:AAA39516.1; PID:g387428
R:Kirschner, L.S.
Nucleic Acids Res. 24, 2829-2834, 1996
A:Title: De novo generation of simple sequence during gene amplification.
A:Reference number: S70711; MUID:96313253
A:Accession: S70711
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 43-92 <KIR>
A:Cross-references: EMBL:U46839; NID:g1228142; PIDN:AAC52722.1; PID:g1228143
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, January 1996
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c
structurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: mdr2
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane
P:1-637, 653-1276/Region: duplication
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:550-554/Region: nucleotide-binding motif B
F:1049-1245/Domain: ATP-binding cassette homology <ABC2>
F:1067-1074/Region: nucleotide-binding motif A (P-loop)
F:1192-1196/Region: nucleotide-binding motif B
F:188,94/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:432/Binding site: ATP (Lys) #status predicted
F:1072/Binding site: ATP (Lys) #status predicted
Query Match 75.2%; Score 4870; DB 1; Length 1276;
Best Local Similarity 74.9%; Pred. No. 2e-273;
Matches 962; Conservative 138; Mismatches 170; Indels 14; Gaps 8;
QY 1 MDPEGGRKGSAPK--NFWKMGKSKK-KEKKEKKPTVSTFAMFRYSNWLDRLYMLVGTMA 57
Db 1 MDLEARNGTARRLDGDFELGSLISNOGREKKKKVNLGILLTIFRYSDMQDKLFMFLGLTLM 60
QY 58 AIIHGAALPLMLLVGNMTDSPA-NAGISRNKTFPVIINESITNNTQHFINHLEEMTY 116
Db 61 AIAHSGSLPLMIVFGEKTDKVDNTG---NFSLPVNFSLSLNPNR---ILEEMTRY 113
QY 117 AYYYSIGAGVILVAAYIQVSWCLAGQIILKIRKQFHAIMRQELGDFVHDVCELNTR 176
Db 114 AYYYSIGGGVILVAAYIQVSWFTLAAGRIKIRKQFFHAILRQEMGWFIDKGTTELNTR 173
QY 177 LTDVSKINEGIDKIGMFFQSIATFFTGFIQVFTGRGKLTILVILAIISPVGLGSAAIWAK 236
Db 174 LTDVSKISEGIDKIGMFFQSIATFFTGFIQVFTGRGKLTILVILAIISPVGLGSAIWA 233
QY 237 ILSSTFDKELLAYAKAGAAVEELAAIRTVIAFGQKKELERYKNKNLEAEKIGIKKAIT 296
Db 234 ILTFSDEKELAAAYAKAGAAVEELAAIRTVIAFGQKKELERYQKHLENAKEIGIKKAI 293
QY 297 ANTSIGAPELLIYASVALAFWYGTSLVLSSEYITIGQVLTVFSSVLIGAFSGQASPSIEA 356
Db 294 ANISMGIAFLLIYASVALAFWYGTSLVLSSEYITIGQVLTVFSSVLIGAFSGQAAAPCIDA 353

Thu Nov 7 09:13:26 2002

Db 772 FLOQTFKAGAILTRLSMAFKAMLRQDMSWEDDYKNSTGALSTRLATDRAQVOGATG 831
Qy 833 SRLAVITONIANLGRGIIISLYGWQLFLLLAIVPIIAIAGVEMKMLSGQALKDKKEL 892
Db 832 TRALIAQNTANLGTGIIISFYGWQLTLLLVVVPFAVSGIVEMKMLAGNAKRDKKAL 891
Qy 893 EGAGKIAIEATENFTVVSLLTREQFEYVYQSLQVYPYNSLRKAHIFGVFSFSTQAMFY 952
Db 892 EAAGKIAIEATENFTVVSLLTREQFEYVYQSLQVYPYNSLRKAHIFGVFSFSTQAMFY 951
Qy 953 FSYAGCFRFGAYLVANEFMNFQDVLVFSIAIVFGAMAYGOVSSFPADYAKAKVSAAHVIM 1012
Db 952 FSYAGCFRFGAYLVANEFMNFQDVLVFSIAIVFGAMAYGOVSSFPADYAKAKVSAAHVIM 1011
Qy 1013 IIEKSPIDSYPHGLKPNLTLEGNTFNEVFNYPYNSLRKAHIFGVFSFSTQAMFY 1072
Db 1012 LFEROPLIDSYPHGLKPNLTLEGNTFNEVFNYPYNSLRKAHIFGVFSFSTQAMFY 1071
Qy 1073 SSGCGKSTVQVLLERFYDPLAGSLVDGKEIKHLNVQWLRHGLGIVSQEPILFDCSIAEN 1132
Db 1072 SSGCGKSTVQVLLERFYDPLAGSLVDGKEIKHLNVQWLRHGLGIVSQEPILFDCSIAEN 1131
Qy 1133 IAYGDNRSVSHHEIVQAAKEANIHFITLPEKYNTRVGDGKTQSLGGQKORIAIARAL 1192
Db 1132 IAYGDNRSVSHHEIVQAAKEANIHFITLPEKYNTRVGDGKTQSLGGQKORIAIARAL 1191
Qy 1193 VROPHILLDEATSDLTSEKVVQALDKAREGRTCIIVIAHRLSTIQNADLIVVFQNGK 1252
Db 1192 VROPHILLDEATSDLTSEKVVQALDKAREGRTCIIVIAHRLSTIQNADLIVVFQNGK 1251
Qy 1253 VKEHGTHQOLLAQKGIYFMSVVOAQAK 1280
Db 1252 VKEHGTHQOLLAQKGIYFMSVVOAQAK 1279

RESULT 10
S41646
p-glycoprotein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 02-Feb-2001
C:Accession: S41646; S22354
R:Brown, P.C.; Thorgeirsson, S.S.; Silverman, J.A.
Nucleic Acids Res. 21, 3885-3891, 1993
A:Title: Cloning and regulation of the rat mdr-2 gene.
A:Reference number: S41646; MUID:93376516
A:Accession: S41646
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1278 <BRO>
A:Cross-references: EMBL:L15079; NID:g310192; PIDN:AAA02937.1; PID:g310193
R:Deuchars, K.L.; Duthie, M.; Ling, V.
Biochim. Biophys. Acta 1130, 157-165, 1992
A:Title: Identification of distinct p-glycoprotein gene sequences in rat.
A:Reference number: S22351; MUID:92223089
A:Accession: S22354
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1211-1278 <DEU>
A:Cross-references: EMBL:X61105; NID:g56892; PIDN:CAA43417.1; PID:g1334220
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:1051-1247/Domain: ATP-binding cassette homology <ABC2>
F:1068-1075/Region: nucleotide-binding motif A (P-loop)

Query Match 74.4% Score 4816; DB 2; Length 1278;
Best Local Similarity 73.9%; Pred. No. 2.6e-270;
Matches 950; Conservative 145; Mismatches 176; Indels 14; Gaps 8;
Qy 1 MDPEGGRGSAEK--NFWKMGK-RSKKKEKKKPTVSTFAMFYSNWDLRLVMTGTA 57
Db 1 MDLEAARNGTARRLDGDFELGSISNQSRKKKKVNLGPIPLTFLFYSDWQDKLFLLGTAM 60

Qy 58 AIIHGAALPLMLVFGNMTDFA-NAGISRNKTFPVIINESITNNNTQHFINHLEEMTY 116
Db 61 AIHAGSLPLMLVFGNMTDFA-NAGISRNKTFPVIINESITNNNTQHFINHLEEMTY 113
Qy 117 AYYSGIGAGVVAAYIQVSWFCLAGROQLTKRQFFHAIMQEQWGFVDVHVDGELNTR 176
Db 114 AYYSGIGAGVVAAYIQVSWFCLAGROQLTKRQFFHAIMQEQWGFVDVHVDGELNTR 173
Qy 177 LTDDSVKINGIGDKIMGFQSIATFTGTFVGTGRGKLTLLVLAISPVLGUSAIAWAK 236
Db 174 LTDDSVKINGIGDKIMGFQSIATFTGTFVGTGRGKLTLLVLAISPVLGUSAIAWAK 233
Qy 237 ILSFTDKELAYAKAGAVAEVLAARTVIAFGGQKKELEYKKNLEAKGIGIKAIT 296
Db 234 ILSFTDKELAYAKAGAVAEVLAARTVIAFGGQKKELEYKKNLEAKGIGIKAIT 293
Qy 297 ANISGAFAFLIYASYALAFWGSTLVISKEYITIGNAMTVFSSILIGAFVSGQAAPCA 356
Db 294 ANISGAFAFLIYASYALAFWGSTLVISKEYITIGNAMTVFSSILIGAFVSGQAAPCA 353
Qy 357 PANARGAAYEIFKIIDNKPISIDYSKSGHKPDNITKNEFNHVSYPSPKVKILKGLN 416
Db 354 PANARGAAYEIFKIIDNKPISIDYSKSGHKPDNITKNEFNHVSYPSPKVKILKGLN 413
Qy 417 LKVGSGQTVALVNGSGCGKSTTVQLMQLYDPTDGMVCIQDIDITINVRHLREITGV 476
Db 414 LKVGSGQTVALVNGSGCGKSTTVQLMQLYDPTDGMVCIQDIDITINVRHLREITGV 473
Qy 477 QEPVLFAITAEINRYGRNVTMDIEIKAVKEANAYDFIMKLPNKFTDLVGERGAQLSGG 536
Db 474 QEPVLFAITAEINRYGRNVTMDIEIKAVKEANAYDFIMKLPNKFTDLVGERGAQLSGG 533
Qy 537 QKQRTAIARALVRNPKILLDEATSDLTSEAEVQALDKAREGRTCIIVIAHRLSTVRN 596
Db 534 QKQRTAIARALVRNPKILLDEATSDLTSEAEVQALDKAREGRTCIIVIAHRLSTVRN 593
Qy 597 ADVIAGFDGIVVEKNGHDELKKEGIYKLVMTQTRGNEIELENATCESKESDALEMS 656
Db 594 ADVIAGFDGIVVEKNGHDELKKEGIYKLVMTQTRGNEIELENATCESKESDALEMS 653
Qy 657 PKDSSSLIKRRSTRRSIHAPQGDRLGKTK-EDLNENVPVSWRILKLNSTWPFV 715
Db 654 P-NGWKARIFRNSTKSLSSRAHQNDVETNELDANVPVFLKVLRLNKTWPFV 712
Qy 716 GTFCAIINGLOPAFISILSILYQWQLTLLAIVPIIAIAGVEMKMLSGQALKDKKELGA 775
Db 713 GTFCAIINGLOPAFISILSILYQWQLTLLAIVPIIAIAGVEMKMLSGQALKDKKELGA 771
Qy 776 GTFCKAGEILTKRLRYMVRSMRLQDYSWFDPKNTTGTALTRLANDAAVQGAIGRL 835
Db 772 GTFCKAGEILTKRLRYMVRSMRLQDYSWFDPKNTTGTALTRLANDAAVQGAIGRL 831
Qy 836 AVITONIANLGRGIIISLYGWQLTLLAIVPIIAIAGVEMKMLSGQALKDKKELGA 895
Db 832 AVITONIANLGRGIIISLYGWQLTLLAIVPIIAIAGVEMKMLSGQALKDKKELGA 891
Qy 896 GKIAIEATENFTVVSLLTREQFEYVYQSLQVYPYNSLRKAHIFGVFSFSTQAMFY 955
Db 892 GKIAIEATENFTVVSLLTREQFEYVYQSLQVYPYNSLRKAHIFGVFSFSTQAMFY 951
Qy 956 AGCFRFGAYLVANEFMNFQDVLVFSIAIVFGAMAYGOVSSFPADYAKAKVSAAHVIM 1015
Db 952 AGCFRFGAYLVANEFMNFQDVLVFSIAIVFGAMAYGOVSSFPADYAKAKVSAAHVIM 1011
Qy 1016 KSPIDSYSPHGLKPNLTLEGNTFNEVFNYPYNSLRKAHIFGVFSFSTQAMFY 1075
Db 1012 KSPIDSYSPHGLKPNLTLEGNTFNEVFNYPYNSLRKAHIFGVFSFSTQAMFY 1071
Qy 1076 CGKSTVQVLLERFYDPLAGSLVDGKEIKHLNVQWLRHGLGIVSQEPILFDCSIAEN 1135
Db 1072 CGKSTVQVLLERFYDPLAGSLVDGKEIKHLNVQWLRHGLGIVSQEPILFDCSIAEN 1131

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
C:Keywords: glycoprotein; membrane protein; nucleotide binding; P-loop
F:455-462/region: nucleotide-binding motif A (P-loop)

Query Match	49.88;	Score 3221.5;	DB 2;	Length 1321;
Best Local Similarity	50.58;	Pred. No. 4.5e-178;		
Matches 650;	Conservative 233;	Mismatches 369;	Indels 35;	Gaps 1
QY	20	KKSKKK-KEKKPTVSTFAMFRYSNWLDRLVMTGMAAIIHGAALPLMLVFGNMTDS	78	
Db	30	KKSRLQDKMKEGDIRVGFELFRSSKDIWMLMGVCALLHGMAGPGLIIFGIMTDI	89	
QY	79	FANAGISR-----NKTFFVINES-----TNTNQHFHLEEMTYAYYSGI	123	
Db	90	FIKYDIERQELIPEKACVNNNTI-VWINSFPHONTNGTCGLVDIESEMKFSGIYAGV	148	
QY	124	GAGVLVAAYIQVSWFCLAGROILKIRQFHAIRMQEIGWFDVHDVGLNTRLTDVSK	183	
Db	149	GMTVLILGYFOIRLWITGARIIRMRKIYFRIRMEIGWFDCTSVGELNSRFADDIEK	208	
QY	184	INEGIGDKIMGFQSTATFTTGFIVGFTGRGKLTIVLAISVPLGSLAAIWAKILSSFTD	243	
Db	209	INDAIADQLAHFLQRMSTAMCGLLGFGYRWKLTIVLAISVPLGIGAAVIGLSIAKTFE	268	
QY	244	KELAYAKAGAAEVLAAIRVIAFGGOKKELERNKLEAKGIGIKKAITANISIGA	303	
Db	269	LELAYAKAGSTADEVLSSIRTVAAFGGENKEVEREYNLVAQRWGWKVMGVMGFTGY	328	
QY	304	AFLLIYASALAFWYGTSLVL-SSEYTGQVLTFFVSLIGAFSGQASPSIEAFANARG	362	
Db	329	MMCLIFCYALAFWYGTSLVLDEEYTPGTLVQIFLCVLAAMNIGHASSCLEIFSTGCS	388	
QY	363	AAEYFIKIDNKPISIDSKSGHKPDNIKGNLFKNVHFSYSRKVEKILKLNKLVQSG	422	
Db	389	AATNIFQTDIQDQVDCMGSDGDKLDRIKEIEFHNVTHYSPRPDKILDLNSWIRKP	448	
QY	423	QTVALVNSGCGKSTTVQLMORLYDPTDGMVCIDGODIRTNVRLHREITGVVSGEPVLV	482	
Db	449	ETVALVSGSGAGKSTALQILQRFYDPCGVMVLDGHDIRSLNIRWLDQIGIVEQEPVLV	508	
QY	483	ATTIAENIRYGRNVMTDEIEKAVKANAYDFIMKLPNKFDTLVGRGAQLSGGQKQRTA	542	
Db	509	STTIAENIRGREDATMEDIVQAADANAYFIMALPQQFDTLVGGGOMSGGQKQVRA	568	
QY	543	IARALVNRKILLDEATSDALDSEAVVQVADLKARKGRTTIVIAHRLSTVRNADVIAG	602	
Db	569	IARALINPKILLDNATSDALDSEARVQEALNKIQHGTIISVAHRLSTVRADVIIG	628	
QY	603	FDDGVIVKGNHDLMEKGIYFKLVTMQTRGNEIELENA-TGESKSESDALEM-----S	656	
Db	629	FEHGVAVERTHEELLERKGVYFMLVTLQSGDNAHKETSIMGKDATEGTLERTFSRGS	688	
QY	657	PKDSGSLIKRRSTR--SIHAP--QGQDKLGTGED-----LNENVPPVSFWIRILKN	706	
Db	689	YRSLRASIRQSRKSQSLTHTDPLAVADHKSSYKDSKDNVDLVVEVEPAPVRILKN	748	
QY	707	STWPPYFWGIFCAIINGLQPAFSIIFSRIGITRDEDPETKRONSNMFSVLFVLGI	766	
Db	749	IPWHYILVGLSAAINGAVTPIYSLFSLQLLGTFSL-LDKEQORSEIHSNCLFFVILGC	807	
QY	767	ISFTTFLOQGTTFKAGEILTKRLRYWFRSMLRDQVSWFDDPKNTTGALTTLANDAAQ	826	
Db	808	VSIFTQLQGTTFKAGEILTKRLKFGFKAMLGQDQIGWFDLNNPVGVLTLRLATDASQ	867	
QY	827	VKGAGSLRVITONTIANLGTGIIISIIYQWLTLLLLAIVPITAIAGVEMKMLSGOAL	886	
Db	868	VQGTGSGVGMVNSFTNIIAALLIAFFFSKLSLIITIFFPFLALSGAVQTKMLTGFAS	927	
QY	887	KDKKELEGAGKIATEAENFTVVSIVLREOKFEYMAQSLQVYRNLSRAHIFGVFSFI	946	
Db	928	QDKQALEKAGQITSEALSNIRTVAGIGVEGRIFKAFVELQTSYKTAVRANIGYLCFAF	987	
QY	947	TOAMMYSYACCFRFGAYLVANFNFQDVLVFSATVFGMAVGQVSSFAPDYAKAKVS	1006	

Db	988	SQIAFLANSAAARYGGYLIAYEGLGFSHVFRVSSVLSATAVGTFTSYTSPSYAKAKIS	1047
QY	1007	AHVIMTIKESPLDSYSPHGLKPNTELEGNVTENEVFNYPTRDPDIPVLQGLSLEVKKGQ	1066
Db	1048	AARFQLLDKRPKPINVYSEAGEKWNDFQKIDFIDCKFTYPSRPIQVLNGLSVSNPQ	1107
QY	1067	TLALVSSGCGKSTVVOLLERYDPLAGSVLDIGKEIKHLNVOWLRAHLGIVSQBPILFD	1126
Db	1108	TLAFVSSGCGKSTSIQLLERYDPOGTVMIDHDSKKVNIQFLRSNIGIVSQBPVLF	1167
QY	1127	CSIAENIAYGNSRVVSHEEIVQAAKEANIHFETLPEKYNTRVYGDGKTQGLSGGQKQRI	1186
Db	1168	CSIMDNIKYDNTKETSVERAIAAAQAOAHDFVMSLPEKYEYTNVGIQSGLSRGEKQRI	1227
QY	1187	ATARALVRPHILLDEATSAIDTSEKVVQEAIDKAREGTCIVIAHRLSTIQNADLIV	1246
Db	1228	ATARAIYRDPKILLDEATSAIDTSEKIVOTALDKAREGTCIVIAHRLSTIQNSDIIA	1287
QY	1247	VFQNGKVEHGT HQOLLAQKGIYFSMV	1273
Db	1288	VVSQGVVIEKGTHEKLMQAKGAYYKLV	1314
RESULT 13			
T42842			
N:Alternate names: bile salt export pump, sister of p-glycoprotein			
C:Species: Rattus norvegicus (Norway rat)			
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-May-2000			
C:Accession: T42842			
R:Gerloff, T.; Steiger, B.; Hagenbuch, B.; Madon, J.; Landmann, L.; Roth, J.; Hofmann			
A:Title: The sister of P-glycoprotein represents the canalicular bile salt export pump			
A:Reference number: 222272; MUID:98212048			
A:Accession: T42842			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: mRNA			
A:Residues: 1-1321 <GR>			
A:Cross-references: EMBL:U69487; NID:g3075421; PID:g3075422; PIDN:AAC40084.1			
A:Experimental source: strain Sprague-Dawley; liver			
C:Genetics:			
A:Gene: spgp			
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog			
C:Keywords: ATP; duplication; glycoprotein; membrane protein; P-loop			
Query Match 49.7%; Score 3217.5; DB 2; Length 1321;			
Best Local Similarity 50.4%; Pred. No. 7.6e-178;			
Matches 649; Conservative 233; Mismatches 370; Indels 35; Gaps 11;			
QY	20	KSKKKK-KEKKPTVSTFAMFRYSNWLDRLVMTGMAAIIHGAALPLMLVFGNMTDS	78
Db	30	KKSLQDKMKEGDIRVGFELFRSSKDIWMLMGVCALLHGMAGPGLIIFGIMTDI	89
QY	79	FANAGISR-----NKTFFVINES-----ITNTQHFHLEEMTYAYYSGI	123
Db	90	FIKYDIERQELIPEKACVNNNTI-VWINSFPHQNTNGTCGLVDIESEMKFSGIYAGV	148
QY	124	GAGVLVAAYIQVSWFCLAGROILKIRQFHAIRMQEIGWFDVHDVGLNTRLTDVSK	183
Db	149	GMTVLILGYFOIRLWITGARQIRMRKIYFRIRMRMEIGWFDCTSVGELNSRFADDIEK	208
QY	184	INEGIGDKIMGFQSIATFTTGFIVGFTGRGKLTIVLAISVPLGSLAAIWAKILSSFTD	243
Db	209	INDAIADQLAHFLQRMSTAMCGLLGFGYRWKLTIVLAISVPLGIGAAVIGLSIAKTFE	268
QY	244	KELAYAKAGAAEVLAAIRVIAFGGOKKELERYKNKLEAKGIGIKKAITANISIGA	303
Db	269	LELAYAKAGSTADEVLSSIRVAFAFGGENKEVEREYNLVAQRWGWKVMGVMGFTGY	328
QY	304	AFLLIYASALAFWYGTSLVL-SSEYTGQVLTFFVSLIGAFSGIQAQSPSIEAFANARG	362
Db	329	MMCLIFFCYALAFWYGTSLVLDEEYTPGTLVQIFLCVLAAMNIGHASSCLEIFSTGCS	388

Db 663 GAPPAPAAEKEIKKKELBEEGAVKANLFKILRYARPEWIIYFFAIIAALIOGAVMPA 722
QY 730 FSIIIESRIIGITRDEDETRKRNMFVSLFVLGLIISFTTFFLQGTFFGKAGEIITKR 789
Db 723 FSLFSQIINVPF-NPDQDKKKGDFWALFVLAAVOGTSMFLQCSLFGVAAERLTMR 781
QY 790 LRYMFRSMLRODSWFDPPDKNTGALTTRLANDAAQVGAIGSELAVITONIANLGTGI 849
Db 782 IRKSVRNVLRODATYFOMPKHSRITRLATDAPNTKSAIDRYGFSINAFISVGGGL 841
QY 850 IISLYGQLTLLAIAIPIAIAAGVEMKMLSGOALKDKKELEGAGKIAATEAENFTV 909
Db 842 GIATYFGOMAFVMAIIPFMAVGQALMKYHGGSATSDAKEMENAGKTAMEAENIRTV 901
QY 910 VSLTREKPEYMAQSLOVYPYRNSLRKAHIFGVFSITQAMMYFSYACGFRGAYLV--A 967
Db 902 QALTQTKLYNIFCSHLDPAGHGNISKATIRGLTYGFANSIQFFTYAAAFRGLFIDFK 961
QY 968 NEMFNQDVLVFSATVFCAMAVGOVSSFPADYAKAKYSAAHVIMIEKSPILDSYSPHG 1027
Db 962 NVLMPEPNVLRVLAISFSFGTIGFAASTYFEIYKATPAAGLIFNMLEEPRIDGWTSSG 1021
QY 1028 LKPNTELEGNVTFNEVYFNFTPRDIPVLQGLSLEVKKGOTLALVSSGCGKSTVVOLLER 1087
Db 1022 TYPO-LSGEVKLNKVFERYPERPAVPILOGLNVHVKPGOTLALVGPSCGKSTVLSLLR 1080
QY 1088 FYPDLAGSLVDKETHLNVQVLRHGLIVGSOEPILEDPCSTAENIAYGDSNRVVSHEEI 1147
Db 1081 LYDPLEGAVTVNDLRQPNKHLRKHIALVSOEPILEDPCSTAENIAYGDSNRVVSHEEI 1140
QY 1148 VQAAKAENIHHFTETLPEKYNTRVGGKQTLGSGQKQRIARAIARVQPHILLDEATSA 1207
Db 1141 ETACSKANIHKFDIDELPDGYETRVGEGKQTLGSGQKQRIARAIARVQPHILLDEATSA 1200
QY 1208 LDTESEKVOQALDKAREGRTCTIVIAHRLSTIONADLIYVVFQNGKVEGHTHQILAAQK 1267
Db 1201 LDTESEKVOQALDKAREGRTCTIVIAHRLSTIONADLIYVVFQNGKVEGHTHQILAAQK 1260
QY 1268 IYFSMWSVQA 1277
Db 1261 AYFALTOKQS 1270

RESULT 15
T31073
multidrug resistance p-glycoprotein - nematode (Haemonchus contortus)
C:Species: Haemonchus contortus
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
R:Xu, M.; Molento, M.; Blackhall, W.; Ribeiro, P.; Beech, R.; Prichard, R.
Mol. Biochem. Parasitol. 91, 327-335, 1998
A:Title: Ivermectin resistance in nematodes may be caused by alteration of P-glycoprotein
A:Reference number: 220968; MUID:98226176
A:Accession: T31073
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1275 <XUM>
A:Cross-references: EMBL:AF003908; NID:g3057041; PID:g3057042; PIDN:AAC36987.1
A:Experimental source: developmental stage adult
C:Genetics:
A:Gene: PGP-A
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology

Query Match 43.6%; Score 2824.5; DB 2; Length 1275;
Best Local Similarity 44.4%; Pred. No. 3.8e-155;
Matches 568; Conservative 271; Mismatches 416; Indels 25; Gaps 12;

QY 8 KGAENKFWKCKSKK-----EKKEKPTVSTFAMFRYSNNLDRVLMVGTMAIIGHA 63
Db 4 KGODDERPLLGKSKSKSGEYSGKEEPTITNNGILSLATLTVLAAAGTAPCVHGA 63
QY 64 ALPLAMLVGNMTDSFANAGISRNKTFPVIINESTTNTHQFINHLEEMTT----YAY 119

Db 64 GFSVLGVLGGMTTVFLRA---QNSEF---VLGTVSRDPEGLPALTKKEEDTIVRRYCLY 117
QY 120 YSGIGAGVLVAAAIQYSFWCCLAAGROILLKIRQOFFHAIMRQIEGNFVDVHVGELNTRLTD 179
Db 118 YLGLGAMFATSIOIQCWETFAERTHKLKIYLAAILRQIQISWFDIQOTGNLTARLTD 177
QY 180 DVSKINIGIGDKTMFQSQSIATFTFTGFTTRGKWLTLVILAIASPVGLSRAIAWAKILS 239
Db 178 DLERVREGLGDKLSLFTQMSAFVAGFCVGSWMTLVMVAVPFIIVISANWMSKIIYA 237
QY 240 SFTDKELLAYAKAGAAVEEVLAAIRTVIAFGGOKKELERYKNKLEAKGIGIKKAITANI 299
Db 238 TRTQVEQETVAGAGIAEETFSIRTVHSCGKHRELTFEEALEKGRQTLGVKVFYMGV 297
QY 300 SIGAAFLIYASALAFWYGTSLVLS--EYTIQOVLTVPFSLVILAFSGIQAQSPSEAPA 358
Db 298 GVGFQMCVTYVYALAFWYGSVLIINDPALDRGRIETVFFAVMSSGAALGCTPLHLNTIS 357
QY 359 NARGAAVEIFKIDNKPSIDSYSKSGHKPNIKGNLFEKNVHFSYPSRKVKILKGLNLK 418
Db 358 IARGAVRSVLVINSRPRKIDPYSLDGIVLNNMRGSIREFKNVHFSYPSRRTLOILKGVSLQ 417
QY 419 VOSGOTVALVNSCGKSTTVQLMORLYDPTDGMVCIDGODIRTIINVHRLREITGVVSQE 478
Db 418 VSQOKTALVSSCGKSTNVNLLRFDYDTRGKVTIDDIQVCLNVQKLEQIGVVSQE 477
QY 479 PVLFTATTIENIRYGRNVMTDEKAEKANAYDFIMKLPNKFDTLVGERGAQLSGQK 538
Db 478 PVLFDGTLFENIKMGYEQATMEEVQEAECRVANAADFTKRLPEGTVTRVGERGVLSGQK 537
QY 539 QRIATARALVNPKILLDEATSAIDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNAD 598
Db 538 QRIATARALVNPKILLDEATSAIDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNAD 597
QY 599 VIAGDDGVIVVEKGNHDELMKEGIYKLVMTQTRGNEIELEENATGESKESDALEMSPK 658
Db 598 QIFVEFKNTVIEQGTTHAELMKNRGVFEEMTQAQVLRQKEEVEILDSDAESDVSPDIALP 657
QY 659 DSGSLIKRSTRRSIHA--POGDRKLGKEDLNENVPVSWFRILKILNSTEWYFVVG 717
Db 658 HLSSRSRKESTRSAISAVPSVRSMQI--EMEDLRKPTPMS--KIFYENRDKWGYFILGL 714
QY 718 FCAIINGGLQPAFSIIFRSIIGITRDEDEPETKRONSMF--SVLFLVLGIISFTFFLQ 776
Db 715 IACIITGVTPTFAVLYAQIIOVYS---EPVDMQKHGVLFWCGAFIVGLVHAFAPFFSA 771
QY 777 FTFKAGEILTRLRYMFRSMLRODSWFDPPDKNTGALTTRLANDAAQVGAIGSRLA 836
Db 772 ICLGRGCEALTKKLRFKFNLLRQNVGVYDIDIRHGTGKLCITRFPATDPNVR--YVFTRLP 830
QY 837 VITONIANLGTGIIISLYGWOLTLALLAIVPIAIAAGVEMKMLSGOALKDKKELEGAG 896
Db 831 GVLSSVVTIIGALVIGFIGWQLALIMVMVPLIIGSGYFEMRMQFGKMRDTELEEAG 890
QY 897 KIATEALENFRVSLTREKPEYMAQSLOVYPYRNSLRKAHIFGVFSITQAMMYFSYA 956
Db 891 KVASQAVENIRTVHALNRQEQEFHFMYCEYLKEPYNELCOAHTYGVFAFSOLIFFMYA 950
QY 957 GCFRFGAYLVANEFMFQDVLVFSIAIVFAGAMAVGOVSSFPADYAKAKYSAAHVIMIEK 1016
Db 951 VAFWIGAIIVDNHISMQPIDYRVFFAFMFCGQWGVNMISSFIPDVVKARLAASILFVLIH 1010
QY 1017 SPLDISYSPHGLKPNLTLEGNTFNEVYFNFTPRDIPVLQGLSLEVKKGOTLALVSSGC 1076
Db 1011 PSEIDNLSDGV--TKKISGHISPRNRYFNPTRRQIRVLRGLNLEINPTGTTVALVGQSGC 1069
QY 1077 GKSTVVOLLERFYDPLAGSVLIDGKEIKHLNVQVLRHGLIVGSOEPILEDPCSTAENIAYG 1136
Db 1070 GKSTVALLERFYDPLAGSVLIDGKEIKHLNVQVLRHGLIVGSOEPILEDPCSTAENIAYG 1129
QY 1137 DNSRVVSHEIIVQAAKEANIHFTIETLPEKYNTRVGGKQTLGSGQKQRIARAIARVROP 1196
Db 1130 LDDPKPSYEQVAAAKMANIHNEVLGLPEGYDTRVGEKGTQSLGSGQKQRIARAIARVROP 1189

QY 1197 HILLDEATSALDTESEKVVQEALDKAREGRTCTIVIAHRLSTIQNADLIWVFONGKVKEH 1256
Db 1190 PILLDEATSALDTESEKIVQDALEVARQRTCTIVIAHRLSTIQSDSVIWMIOEGKATDR 1249
QY 1257 GTHOOLLAQKGIYFSMVSQ 1276
Db 1250 GTHELLMKNDLYKRLCETQ 1269

Search completed: November 6, 2002, 18:45:05
Job time : 28.063 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:30:44 ; Search time 10.4981 Seconds
(without alignments)
4724.646 Million cell updates/sec

Title: US-09-672-725C-27

Perfect score: 6473

Sequence: 1 MDPEGRKGSAEKFWKMGK.....LLAQKGIYFSMVSVQAGAKR 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5866	90.6	1280	1	MDR1_HUMAN
2	5664.5	87.5	1276	1	MDR1_CRIGR
3	5638	87.1	1276	1	MDR3_MOUSE
4	5301	81.9	1276	1	MDR1_MOUSE
5	5280	81.6	1276	1	MDR2_CRIGR
6	5231.5	80.8	1277	1	MDR1_RAT
7	4922.5	76.0	1279	1	MDR3_HUMAN
8	4870	75.2	1276	1	MDR2_MOUSE
9	4829.5	74.6	1281	1	MDR3_CRIGR
10	4816	74.4	1278	1	MDR2_RAT
11	3308.5	51.1	1321	1	AB11_HUMAN
12	3229.5	49.9	1321	1	AB11_RABIT
13	3219.5	49.7	1321	1	AB11_MOUSE
14	3217.5	49.7	1321	1	AB11_RAT
15	2741	42.3	1321	1	MDR1_CAEEL
16	2568.5	39.7	1302	1	MDR4_DROME
17	2497	38.6	1302	1	MDR5_DROME
18	2338.5	36.1	1254	1	MDR3_CAEEL
19	2276	35.2	1362	1	PMD1_SCHPO
20	2074	32.0	1280	1	MDR1_LEJEN
21	1544.5	23.9	1419	1	MDR_PLAFF
22	1219.5	18.8	1290	1	STE5_YEAST
23	1122	17.3	1336	1	MAM1_SCHPO
24	1030.5	15.9	1323	1	HST6_CANAL
25	892	13.8	735	1	ABC8_HUMAN
26	827	12.8	582	1	MSBA_ECOLI
27	814	12.6	587	1	MSBA_HAEIN
28	768	11.9	598	1	Y288_THEMA
29	767	11.8	820	1	MDL2_YEAST
30	765	11.8	695	1	MDL1_YEAST
31	761	11.7	1437	1	MRP5_HUMAN
32	755.5	11.7	1436	1	MRP5_MOUSE
33	745.5	11.5	1436	1	MRP5_MOUSE

34	744.5	11.5	726	1	YFX9_SCHPO
35	741	11.4	1545	1	MRP2_HUMAN
36	730	11.3	1564	1	MRP2_RABIT
37	728	11.2	1325	1	MRP4_HUMAN
38	723.5	11.2	703	1	TAP2_RAT
39	721	11.1	685	1	MDL1_CANAL
40	708	10.9	1541	1	MRP2_RAT
41	706	10.9	584	1	LMRA_LACLA
42	706	10.9	604	1	YFIC_BACSU
43	703.5	10.9	607	1	HEPA_ANASP
44	703	10.9	575	1	YWJA_BNCSU
45	694.5	10.7	702	1	TAP2_MOUSE

ALIGNMENTS

RESULT 1

MDR1_HUMAN					
ID	MDR1_HUMAN	STANDARD;	PRT; 1280 AA.		
AC	P08183; Q12755; Q14812;				
DT	01-AUG-1988 (Rel. 08, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Multidrug resistance protein 1 (P-glycoprotein 1).				
GN	ABCBI OR PGI1 OR MDR1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=87028230; PubMed=2876781;				
RA	Chen C.-J., Chin J.E., Ueda K., Clark D.P., Pastan I., Gottesman M.M.,				
RA	Roninson I.B.;				
RT	"Internal duplication and homology with bacterial transport proteins				
RT	in the mdr1 (P-glycoprotein) gene from multidrug-resistant human				
RT	cells.";				
RL	Cell 47:381-389(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90094448; PubMed=1967175;				
RA	Chen C.-J., Clark D.P., Ueda K., Pastan I., Gottesman M.M.,				
RA	Roninson I.B.;				
RT	"Genomic organization of the human multidrug resistance (MDR1) gene				
RT	and origin of P-glycoproteins.";				
RT	J. Biol. Chem. 265:506-514(1990).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97190336; PubMed=9038218;				
RA	Chen G., Duran G.E., Steger K.A., Lacayo N.J., Jaffrezou J.P.,				
RA	Dumontet C., Sikic B.I.;				
RT	"Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein,				
RT	altered phenotype, and resistance to cyclosporins.";				
RT	J. Biol. Chem. 272:5974-5982(1997).				
RN	[4]				
RP	SEQUENCE OF 1-234 FROM N.A.				
RA	Smith A., Beck C., Gibson A.;				
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.				
RN	[5]				
RP	SEQUENCE OF 178-215 AND 800-856 FROM N.A.				
RX	MEDLINE=90290529; PubMed=1972623;				
RA	Gekeler V., Weger S., Probst H.;				
RT	"mdr1P-glycoprotein gene segments analyzed from various human				
RT	leukemic cell lines exhibiting different multidrug resistance				
RT	profiles.";				
RL	Biochem. Biophys. Res. Commun. 169:796-802(1990).				
RN	[6]				
RP	SEQUENCE OF 1-23 FROM N.A.				
RA	Kioka N., Tsubota J., Kakehi Y., Komano T., Gottesman M.M.,				
RA	Pastan I., Uedak K.;				
RL	Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.				
CC	!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED				

Q9Y7m7	schizosacch
Q92887	homo sapien
Q28689	oryctolagus
O15439	homo sapien
P36372	rattus norv
Q63120	candida alb
Q9chl8	lactococcus
P54719	bacillus su
P22638	anaeana su
P45861	bacillus su
P36371	mus musculus

CC CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, SMALL INTESTINE
CC CC AND BRAIN
CC CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC CC or send an email to license@isb-sib.ch).
CC CC -----

CC EMBL: M14758; AAA59575.1; -
CC EMBL: M29447; AAA59576.1; -
CC EMBL: M29424; AAA59576.1; JOINED.
CC EMBL: M29425; AAA59576.1; JOINED.
CC EMBL: M29426; AAA59576.1; JOINED.
CC EMBL: M29427; AAA59576.1; JOINED.
CC EMBL: M29428; AAA59576.1; JOINED.
CC EMBL: M29429; AAA59576.1; JOINED.
CC EMBL: M29430; AAA59576.1; JOINED.
CC EMBL: M29431; AAA59576.1; JOINED.
CC EMBL: M29432; AAA59576.1; JOINED.
CC EMBL: M29433; AAA59576.1; JOINED.
CC EMBL: M29434; AAA59576.1; JOINED.
CC EMBL: M29435; AAA59576.1; JOINED.
CC EMBL: M29436; AAA59576.1; JOINED.
CC EMBL: M29437; AAA59576.1; JOINED.
CC EMBL: M29438; AAA59576.1; JOINED.
CC EMBL: M29439; AAA59576.1; JOINED.
CC EMBL: M29440; AAA59576.1; JOINED.
CC EMBL: M29441; AAA59576.1; JOINED.
CC EMBL: M29442; AAA59576.1; JOINED.
CC EMBL: M29443; AAA59576.1; JOINED.
CC EMBL: M29444; AAA59576.1; JOINED.
CC EMBL: M29445; AAA59576.1; JOINED.
CC EMBL: M29446; AAA59576.1; JOINED.
CC EMBL: AF016535; AAB69423.1; -
CC EMBL: AC002457; AAC82531.1; -
CC EMBL: M37724; AAA88047.1; -
CC EMBL: M37725; AAA88048.1; -
CC EMBL: X58723; CAA41558.1; -
CC PIR: A25059; DVHUI.
CC PIR: A34914; A34914.
CC MIM: 171050; -
CC InterPro: IPR003593; AAA.
CC InterPro: IPR001140; ABC_transporter_tmemb.
CC InterPro: IPR003439; ABC_transporter.
CC InterPro: IPR001687; ATP_GTP_A.
CC Pfam: PF00664; ABC_membrane; 2.
CC Pfam: PF00005; ABC_tran; 2.
CC SMART: SM00382; AAA; 2.
CC PROSITE: PS00211; ABC_TRANSPORTER; 2.
CC ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
CC Multigene family.
CC DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 52 72 POTENTIAL.
CC TRANSMEM 120 140 POTENTIAL.
CC TRANSMEM 189 209 POTENTIAL.
CC TRANSMEM 216 236 POTENTIAL.
CC TRANSMEM 297 317 POTENTIAL.
CC TRANSMEM 326 346 POTENTIAL.
CC DOMAIN 347 710 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 711 731 POTENTIAL.
CC TRANSMEM 757 777 POTENTIAL.
CC TRANSMEM 833 853 POTENTIAL.
CC TRANSMEM 854 874 POTENTIAL.
CC TRANSMEM 937 957 POTENTIAL.
CC TRANSMEM 974 994 POTENTIAL.
CC DOMAIN 995 1280 CYTOPLASMIC (POTENTIAL).
CC NP_BIND 427 434 ATP (BY SIMILARITY).

FT NP_BIND 1070 1077 ATP (BY SIMILARITY).
FT REPEAT 1 637
FT REPEAT 638 1280
FT CARBOHYD 91
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 23 23 S -> R (IN REF. 6).
FT CONFLICT 185 185 G -> V (IN REF. 1 AND 3).
FT CONFLICT 336 336 MISSING (IN REF. 3).
FT CONFLICT 412 412 G -> A (IN REF. 3).
FT CONFLICT 438 438 Q -> S (IN REF. 3).
SQ SEQUENCE 1280 AA; 141462 MW; 141462 MW; ALC279531F43675 CRC64;

Query Match 90.6%; Score 5866; DB 1; Length 1280;
Best Local Similarity 90.8%; Pred. No. 0;
Matches 1164; Conservative 55; Mismatches 59; Indels 4; Gaps 4;

QY 1 MDPGGRKGSA-EKNFWKMGKKKKKKPTVSTFAMFRYSNWLDRMLVGMVMAAI 59
Db 1 MDLEGDRGGAKKNFKLNKNS-EKDKKKPTVSTFAMFRYSNWLDRMLVGMVMAAI 59

QY 60 IHGAALPLMLVFCNMTDSFANAGISRNKTFPPVINESITNTQHFTHNHEEEMTVYAY 119
Db 60 IHGAGLPLMLVFCNMTDSFANAGISRNKTFPPVINESITNTQHFTHNHEEEMTVYAY 117

QY 120 YSGIGAGVLVAAYTOVSFWCLAAAGQILKIRKQFFHAIMRQEGWFDYHDVGEINLTLD 179
Db 118 YSGIGAGVLVAAYTOVSFWCLAAAGQILKIRKQFFHAIMRQEGWFDYHDVGEINLTLD 177

QY 180 DYSKINEGIDKIGMFQSIATFETGFTVGRGKTLTLVLAISPVGLSAAIAWAKILS 239
Db 178 DYSKINEGIDKIGMFQSIATFETGFTVGRGKTLTLVLAISPVGLSAAIAWAKILS 237

QY 240 SFTDKELLAYAKAGAAVEVLAAIRTVIAFGGQKKELRYNKNLEAKGIGIKKAITANI 299
Db 238 SFTDKELLAYAKAGAAVEVLAAIRTVIAFGGQKKELRYNKNLEAKGIGIKKAITANI 297

QY 300 STGAFLLIYAYALAFWYGTSLVLSSEYTTGOVLTVFSSVLIGAFSGQASPSIEAFAN 359
Db 298 STGAFLLIYAYALAFWYGTSLVLSSEYTTGOVLTVFSSVLIGAFSGQASPSIEAFAN 357

QY 360 ARGAAYEIFKIIDNKPISIDSKSGHKPDNKGNEFNHVSFSPSRKEVKILKGLNLKV 419
Db 358 ARGAAYEIFKIIDNKPISIDSKSGHKPDNKGNEFNHVSFSPSRKEVKILKGLNLKV 417

QY 420 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTGEMVSDGQDINTINVRFLREITGVVSOEP 479
Db 418 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTGEMVSDGQDINTINVRFLREITGVVSOEP 477

QY 480 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKQ 539
Db 478 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKQ 537

QY 540 RTAIALRVNPKILLDLDEATSDALDTESEAVVQVADLKARKGRTTIVIAHRLSTVRNADV 599
Db 538 RTAIALRVNPKILLDLDEATSDALDTESEAVVQVADLKARKGRTTIVIAHRLSTVRNADV 597

QY 600 IAGFDGVIKGNHDELKMKGIYFKLVTMOTRGNIELENAETGESDESALDMSKPD 659
Db 598 IAGFDGVIKGNHDELKMKGIYFKLVTMOTRGNIELENAETGESDESALDMSKPD 657

QY 660 SGSSLLKRRSTRRSIHAPOQQRKLGTKEDLNENPPVSWFRILKLNSTWEPYFVVGIFC 719
Db 658 SRSSLLKRRSTRRSIHAPOQQRKLGTKEDLNENPPVSWFRILKLNSTWEPYFVVGIFC 717

QY 720 AINGGLQPAFSTIFRSRIIGITFRDDEPDKRQNSWFSVLVLGLIISFITFELQGF 779
Db 718 AINGGLQPAFSTIFRSRIIGITFRDDEPDKRQNSWFSVLVLGLIISFITFELQGF 777

QY 780 GKAGEILTKRLRYMVFRLRQDVSWFDDPKNTTGALTTRLANDAAQVKGAGSLRAVIT 839
Db 778 GKAGEILTKRLRYMVFRLRQDVSWFDDPKNTTGALTTRLANDAAQVKGAGSLRAVIT 837


```

Db 295 SGAFAFLLIYALAFWGTSLVTSKEYSGQVTVFFAFLIGAFSLGQSPNIEAFAN 354
Qy 360 ARGAAEYEFKIIDNKPSIDSYSGKHGPDNIKGNLEFKNVHFSPSRKEVKILKGLNLKV 419
Db 355 ARGAAEYEFNIIDNKPSIDFSKNGYKPDNIKGNLEFKNVHFSPSRKDVQLKGLNLKV 414
Qy 420 QSGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMWCIDGQDRTINVRHLREITGVVSOEP 479
Db 415 QSGQTVLVGNSGCGKSTTVQLQLRLYDPTGEGVSDGQDRTINVRYLREITGVVSOEP 474
Qy 480 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKQ 539
Db 475 VLFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKQ 534
Qy 540 RTAARALVRNPKILLDEATSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADV 599
Db 535 RTAARALVRNPKILLDEATSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADI 594
Qy 600 IAGFDGCVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKSEDALEWSPKD 659
Db 595 IAGFDGCVIVEGKNHDELMKEGIYFKLVMTQTAGNEIELGNEVEGESKEIDNLDMSKD 654
Qy 660 SGSSLIKRRSTRRSIAPQOGQRKLTGEDLNENPVVPSFWIRILKLNSTEWPFYVVGIFC 719
Db 655 SASLLRRSTRRSIRGPHDQDKLSTKEALDEDVPPISFWIRILKLNSENWPFYVVGIFC 714
Qy 720 AINGGLQPAFSIIFRIIGITFRDEPDKRQNSNMVSLFVLGILGISFTIFFLQGTFF 779
Db 715 AIVNGALQPAFSIIFSKVGVFTRNTDDETKRHDSNLFSLFLILGIVSFTIFFLQGTFF 774
Qy 780 GKAGEILTKRLRYWPKSMURQDVSFDPKNTGALTTRLANDAOVKGATGSLRAVIT 839
Db 775 GKAGEILTKRLRYWPKSMURQDVSFDPKNTGALTTRLANDACQVKGATGARLAVIT 834
Qy 840 QNIANLGTGIIISLIYGWQLTLLLAIVPIIAIAGVVMKMLSGQALKDKKELEGAKTA 899
Db 835 QNIANLGTGIIISLIYGWQLTLLLAIVPIIAIAGVVMKMLSGQALKDKKELEGSKTA 894
Qy 900 TEAENPRTVSLTREOKFEYMAQSLQVYPRNSLRKAHFGVFSFISITQAMVFSVAGCF 959
Db 895 TEAENPRTVSLTREOKFENYMAQSLQVYPRNALKKAHVFGITFTQAMVFSVAGCF 954
Qy 960 RFGAYLVANEFMFQDVLVLSAIVFGAMAVGVQVSSFADPYAKAKVSAAHVIMIEKSP 1019
Db 955 RFGAYLVARELMTFENVLLVLSAIVFGAMAVGVQVSSFADPYAKAKVSASHIIMIEKVP 1014
Qy 1020 IDSVSPHGLKPNTELEGNVTNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVGS CGKS 1079
Db 1015 IDSVSTGGLKPNTELEGNVFNFNYPTRPDIPVLQGLNLEVKKGQTLALVGS CGKS 1074
Qy 1080 TVVOLLERFDYDPLAGSVLIDGKEIKHLNWQLRAHGLGIVSOEPIFDCSTAEINAYGDNS 1139
Db 1075 TVVOLLERFDYDPMAGTVFDLQGVNQVNLQWLRHAHGLGIVSOEPIFDCSTAEINAYGDNS 1134
Qy 1140 RVVSHIEIVQAAKENIHFETLPEKYNTRVGDKGTQSLGGQKQRTATARALVRPHIL 1199
Db 1135 RVVQDEIERAAKENIHFQIESLPDKYNTRVGDKGTQSLGGQKQRTATARALVRPHIL 1194
Qy 1200 LLDATSEKLVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1259
Db 1195 LLDATSEKLVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1254
Qy 1260 QOLLAQKGIYFSMVSVQAGAKR 1281
Db 1255 QOLLAQKGIYFSMVSVQAGAKR 1276
RESULT 3
MDR3_MOUSE STANDARD; PRT; 1276 AA.
AC P21447;
DT 01-MAY-1991 (Rel. 18, Created)
```

```

DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (P-glycoprotein 3) (MDR1A).
GN ABCB4 OR PGY3 OR PGY-3 OR MDR3 OR MDR1A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205845; PubMed=1969610;
RA Devault A., Gros P.;
RT "Two members of the mouse mdr gene family confer multidrug resistance
RL with overlapping but distinct drug specificities.";
RL Mol. Cell. Biol. 10:1652-1663(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90287150; PubMed=1972547;
RA Hsu S.I.H., Cohen D., Kirschner L.S., Lothstein L., Hartstein M.,
RA Horwitz S.B.;
RT "Structural analysis of the mouse mdr1a (P-glycoprotein) promoter
RT reveals the basis for differential transcript heterogeneity in
RT multidrug-resistant J774.2 cells.";
RL Mol. Cell. Biol. 10:3596-3608(1990).
RN [3]
RP SEQUENCE OF 173-1276 FROM N.A.
RX STRAIN=BALB/C;
RX MEDLINE=89308614; PubMed=2473069;
RA Hsu S.I.H., Lothstein L., Horwitz S.B.;
RT "Differential overexpression of three mdr gene family members in
RT multidrug-resistant J774.2 mouse cells. Evidence that distinct P-
RT glycoprotein precursors are encoded by unique mdr genes.";
RL J. Biol. Chem. 264:12053-12062(1989).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
CC RELATED BUT DISTINCT CELLULAR GENES.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M30697; AAA39517.1; -
CC EMBL; M33581; AAA39514.1; -
CC EMBL; M33580; AAA39518.1; -
CC EMBL; M24417; AAA03243.1; -
CC PIR; A34175; DVM5LA.
CC PIR; A34786; A34786.
CC MGD; MGI:97570; Abcb4.
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001140; ABC_transporter_tmem.
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR001687; ATP_GTP_A.
CC Pfam; PF00664; ABC_membrane; 2.
CC Pfam; PF00005; ABC_tran; 2.
CC SMART; SM00382; AAA; 2.
CC PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 48 71 POTENTIAL.
FT TRANSMEM 117 136 POTENTIAL.
FT TRANSMEM 188 205 POTENTIAL.
FT TRANSMEM 212 232 POTENTIAL.
FT TRANSMEM 298 318 POTENTIAL.
FT TRANSMEM 327 346 POTENTIAL.
FT DOMAIN 347 707 CYTOPLASMIC (POTENTIAL).
```

```

FT TRANSMEM 708 728 POTENTIAL.
FT TRANSMEM 753 773 POTENTIAL.
FT TRANSMEM 829 848 POTENTIAL.
FT TRANSMEM 853 880 POTENTIAL.
FT TRANSMEM 942 961 POTENTIAL.
FT TRANSMEM 964 984 POTENTIAL.
FT DOMAIN 985 1276 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 423 430 ATP (BY SIMILARITY).
FT NP_BIND 1066 1073 ATP (BY SIMILARITY).
FT REPEAT 1 632
FT REPEAT 633 1276
FT CARBOHYD 83
FT CARBOHYD 87 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 526 527 QL -> HV (IN REF. 2 AND 3).
FT CONFLICT 939 939 F -> S (IN REF. 2 AND 3).
FT CONFLICT 1036 1036 F -> V (IN REF. 2 AND 3).
SQ SEQUENCE 1276 AA; 140754 MW; 75C71F3E1F58481 CRC64;

Query Match 87.1%; Score 5638; DB 1; Length 1276;
Best Local Similarity 87.1%; Pred. No. 2.2e-314;
Matches 1116; Conservative 74; Mismatches 84; Indels 8; Gaps 4;

QY 1 MDPEGRKSGAEKFNKMGKKKKKKKKKPTVSTFAMPRYSNWLDRMLYMLVGTMAAII 60
DB 1 MELEEDLKGRAKNFSGKMGKKS-KKEKKKKPAVSVLTMFRYAGWLDRLYMLVGTMAAII 59

QY 61 HGAALPLMLVFGNMVTDSPANAG-ISRNTKTFPVIIINESITNNQTHFINHLEEMTYIAY 119
DB 60 HGVALPLMLVFGNMVTDSPANAGVGNVSKNST-----NMSEADKRAMFAK-LEEEMTYIAY 113

QY 120 YSGTGAGVLVAAYIQVSWFCLAAAGRLKIRKQFFHAIMRQETGDFVDHVDGELNRLTD 179
DB 114 YTGIGAGVLIVAYIQVSWFCLAAAGROIHKIRQKFFHAIMNQELGDFVDHVDGELNRLTD 173

QY 180 DVSKINEGIDGKGMFQSTATFTTGIVGFTRGWKLTVILAIAPVLGSLAAINWAKILS 239
DB 174 DVSKINEGIDGKGMFQAMATFFGFIIGFTRGWKLTVILAIAPVLGSLAGINWAKILS 233

QY 240 SFTDKELLAYAKAGAAVEEVLAIIRTVIAFGGOKKELERYKNLEAKGIGIKKAITANI 299
DB 234 SFTDKELHAYAKAGAAVEEVLAIIRTVIAFGGOKKELERYNNLEAKRLGKIKKAITANI 293

QY 300 SIGAAPELLIYASALAFWYGTSLVLSSEYTIQVLTVPFVSVLIGAFSIGQASPSIEAFAN 359
DB 294 SMGAPELLIYASALAFWYGTSLVLSKEYSIGQVLTVPFVSVLIGAFSVQASPNIEAFAN 353

QY 360 ARGAAYEIEKIIDNKPISDSYSGHKKPDNIKNLEFNKFNHSPSRKEVKILKGLNLKV 419
DB 354 ARGAAYEVEFKIIDNKPISDSYSGHKKPDNIQGNLEFNKFNHSPSRKEVQILKGLNLKV 413

QY 420 QSGQTVALVGNSGCGKSTTVQMLMORLYDPTDGMVCIDGDIIRTNVRLHRELTGVVVSQEP 479
DB 414 KSGQTVALVGNSGCGKSTTVQMLMORLYDPLDGMVSDGDIIRTNVRLHRELTGVVVSQEP 473

QY 480 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKQ 539
DB 474 VLFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGGQKQ 533

QY 540 RIAIARALVRNPKILLLDEATSALETSEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 599
DB 534 RIAIARALVRNPKILLLDEATSALETSEAVVQVALDKARKGRTTIVIAHRLSTVRNADV 593

QY 600 IAGFDGDIVVEKGNHDELMKEGIYFKLVMTQTRNGNELEENATGESKESDALRMSPKD 659
DB 594 IAGFDGDIVVEKGNHDELMKEGIYFKLVMTQTRNGNELEENATGESKESDALRMSPKD 653

QY 660 SGSSLIKRSTRSRTHAPOGQDRKLTGKEDLNENYPPVSFWRLKLNSTWPFYVVGIFC 719
DB 654 SGSSLIKRSTRSRTHAPOGQDRKLTGKEDLNENYPPVSFWRLKLNSTWPFYVVGIFC 713

QY 720 AIINGLOPAFSAIIFSRIGTFRDEDETPKQNSMFSVFLVIGLISFTIFFLQGGTF 779
DB 714 AIINGLOPAFSAIIFSKVVGVTNGGPPETQRONSNLFSLLFLILGIISFTIFFLQGGTF 773
DB 780 GKAGEILTKRLRYMYFRSMRLRQDVSWFDDPKNTTGALTRLANDAAQVKGATGSLAVIT 839
DB 774 GKAGEILTKRLRYMYFRSMRLRQDVSWFDDPKNTTGALTRLANDAAQVKGATGSLAVIT 833
QY 840 QNTANLGTGIIISLIYWGWLTLTLLLAIVPIIAIAGVVEKMLSGQALKDKKELEGAGKIA 899
DB 834 QNTANLGTGIIISLIYWGWLTLTLLLAIVPIIAIAGVVEKMLSGQALKDKKELEGAGKIA 893
QY 900 TEAIENFRVSLTROKFEYMAQSLQVPRNSLRKAHIFGVFSFISITQAMMYFSYACGF 959
DB 894 TEAIENFRVSLTROKFEYMAQSLQVPRNSLRKAHIFGVFSFISITQAMMYFSYACGF 953
QY 960 RFGAYLVANEFNMFDVLLVFSIAIVFGAMAVQGVSSFPADYAKAKVSAAHVIMITEKSPL 1019
DB 954 RFGAYLVTOQLMTFENVLLVFSIAIVFGAMAVQGVSSFPADYAKATVSASHIIRIEKTP 1013
QY 1020 IDSYSPLHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLOGLSLEVKKGTALALVSSGCGKS 1079
DB 1014 IDSYSTQGLKPNLTLEGNVTFNEVFNYPTRPDIPVLOGLSLEVKKGTALALVSSGCGKS 1073
QY 1080 TVVQLLEFYDPLAGSVLIDGKEIKHLNVQWLRHLGIVSQEPILFDCSIAENIAYGNS 1139
DB 1074 TVVQLLEFYDPMAGSVLIDGKEIKHLNVQWLRHLGIVSQEPILFDCSIAENIAYGNS 1133
QY 1140 RVSHHEIVQAAKENIHHFIETLPEKYNTRYVGDGKTOLSGGQKORIAIARALVRQPHIL 1199
DB 1134 RVSYEEIVRAAKENIHHFIETLPEKYNTRYVGDGKTOLSGGQKORIAIARALVRQPHIL 1193
QY 1200 LDEATSALETSEKVVQVALDKARKGRTTIVIAHRLSTIONADLVVFGKVEKHGTH 1259
DB 1194 LDEATSALETSEKVVQVALDKARKGRTTIVIAHRLSTIONADLVVFGKVEKHGTH 1253
QY 1260 QQLLAQKGIYFSWVSQVQAKR 1281
DB 1254 QQLLAQKGIYFSWVSQVQAKR 1275

RESULT 4
MDRL_MOUSE STANDARD; PRT; 1276 AA.
AC P06795;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein 1).
GN ABCB1 OR PGY1 OR PGY1-1 OR MDRL1 OR MDRLB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87028229; PubMed=3768958;
RA Gros P., Croop J., Housman D.;
RT "Mammalian multidrug resistance gene: complete cDNA sequence
RT indicates strong homology to bacterial transport proteins.";
RL Cell 47:371-380(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89367274; PubMed=2570420;
RA Raymond M., Gros P.;
RT "Mammalian multidrug-resistance gene: correlation of exon
RT organization with structural domains and duplication of an ancestral
RT gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:6488-6492(1989).
RN [3]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=91042535; PubMed=2248681;
RA Raymond M., Gros P.;
RT "Cell-specific activity of cis-acting regulatory elements in the
RT promoter of the mouse multidrug resistance gene mdrl.";

```

RL Mol. Cell. Biol. 10:6036-6040(1990).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- PTM: SEVERAL PHOSPHORYLATED SERINE RESIDUES ARE PRESENT IN THE
CC LINKER DOMAIN.
CC -!- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
CC RELATED BUT DISTINCT CELLULAR GENES.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M14757; AAA79005.1; -.
DR EMBL: M60348; AAA39513.1; -.
DR PIR: A33719; DVMSL.
DR MGI: MGI:97568; Abcb1.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC_transporter_tmern.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane; 2.
DR Pfam: PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family; Phosphorylation.
FT DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 48 71 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 327 347 POTENTIAL.
FT DOMAIN 348 709 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 710 730 POTENTIAL.
FT TRANSMEM 755 775 POTENTIAL.
FT TRANSMEM 831 851 POTENTIAL.
FT TRANSMEM 855 882 POTENTIAL.
FT TRANSMEM 935 955 POTENTIAL.
FT TRANSMEM 966 986 POTENTIAL.
FT DOMAIN 987 1276 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (BY SIMILARITY).
FT NP_BIND 1068 1075 ATP (BY SIMILARITY).
FT REPEAT 1 635
FT REPEAT 636 1276
FT CARBOHYD 73 73 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1276 AA; 140993 MW; 18040F011B0FF4E CRC64;
Query Match 81.9%; Score 5301; DB 1; Length 1276;
Best Local Similarity 80.3%; Pred. No. 3.9e-295;
Matches 1030; Conservative 130; Mismatches 114; Indels 8; Gaps 5;
Qy 1 MDPEGGKRGSAEKNFWMGKSKKKEKKKPTVTFAMFRYSNWLDRMLVGTAAII 60
Db 1 MEFEENLKGADKNFSGMKGKS-KKEKKKPAVGVGFMFRYADWLDKLMLGTAAII 59
Qy 61 HGAALPLMLVFGNWTDFANAGTSRNKTPVIINESITNTQHFIN-HLEEMETVAYY 119
Db 60 HGTLLPLMLVFGNWTDFSTKAES---ILPSITNQSNGPNTLIISNLSLEEMAIYAY 116
Qy 120 YSGIGAGLVAAIQVSWFLAAGRQILKIRKOFFHAIMRQEGWFDVHDVGEINTRLTD 179
Db 117 YTGIGAGLVIAIYQVSWFLAAGRQILKIRKOFFHAIMRQEGWFDVHDVGEINTRLTD 176
Qy 180 DVSKINEGIGDKIGMFPQSIATFTTGTGVTGTRGKTLTLVLAISPVGLSAAIWAII 239

Db 177 DVSKINDIGDKGMFPQSIATFTTLAGPIIGFISGKLTLLVLAIVSPGLISGLSAAKAVLT 236
Qy 240 SFTDKELLAYAKAGAAVEEVLAAIRTVIAFGGQKKELERYKNKLEAKGIGIKKAITANI 299
Db 237 SFTNKELQAKAGAAVEEVLAAIRTVIAFGGQKKELERYKNKLEAKGIVGKKAITASI 296
Qy 300 SIGAAFLIYASALAFWYGTSLVLSSEYTIQVLTIVFFSVLIGAFSGIQAASPIAFAN 359
Db 297 SIGIAYLLVYASALAFWYGTSLVLSSEYTIQVLTIVFFSVLIGAFSGIQAASPIAFAN 356
Qy 360 ARGAAEYFIKIIDNKPSIDSKSGHPPDNKGNLEPKNVHSPSKVEKILKGLNLKV 419
Db 357 ARGAAEYFIKIIDNEPIDSFSTKGYKPDSTMGNLEPKNVHFPSPSEVQILKGLNLKV 416
Qy 420 QSGQTVLALVNGSGCGKSTTVQLMQRLYDPTDGMVCDIGQDIRTINVRHLREITGVWSQEP 479
Db 417 KSGQTVLALVNGSGCGKSTTVQLMQRLYDPTDGMVCDIGQDIRTINVRHLREITGVWSQEP 476
Qy 480 VLFATTIAENIRYGRENVTMDEIEKAVKEANAYDFIMKLPNKFTDVLGERGAQLSGQKQ 539
Db 477 VLFATTIAENIRYGRENVTMDEIEKAVKEANAYDFIMKLPNKFTDVLGERGAQLSGQKQ 536
Qy 540 RIATARALVRNPKILLDEATSLDTSSEAVVQVALDKARKGRITIVIAHRLSTVRNADV 599
Db 537 RIATARALVRNPKILLDEATSLDTSSEAVVQVALDKARKGRITIVIAHRLSTVRNADV 596
Qy 600 IAGFDGQVIVEKGNHDELMKEGIYFKLVMTQTRNGEIELENATGESKSSDALEMPK 659
Db 597 IAGFDGQVIVEKGNHDELMKEGIYFKLVMTQTRNGEIELENATGESKSSDALEMPK 656
Qy 660 SGSLIKRRSTRSRSIHAPQGGDKLTKEDLNENVPVSWRILKLNSTWPEYVVGIFC 719
Db 657 SKSPIL-RRSIYRSVHRKQDQERLSMKEADEVDPVLSFWRLNLNLSEWPLYLVGVLC 715
Qy 720 AINGLQPAFSIIFRIIGITRDEDPETKRQNSNFVSLFLVGLISITFTFLOGFTF 779
Db 716 AVINGCIQPAIVFSRIVGVFSRDDHETKRQCNLFSLFLVGLISITFTFLOGFTF 775
Qy 780 GKAGEILTKRLRYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAQAQGAIGSLAVIT 839
Db 776 GKAGEILTKRLRYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAQAQGAIGSLAVIT 835
Qy 840 QNIANLGTGIIISLYGWQLTLIIAIVPIIAIAGVVMKMLSGOALKDKKLEGAQKIA 899
Db 836 QNVANLGTGIVLSLYGWQLTLIIAIVPIIAIAGVVMKMLSGOALKDKKLEGAQKIA 895
Qy 900 TEAIENFRTVSLTREQKFEYMAQSLQVPYRNSLRKAHIFGVFSFSTIQAMMYFSYAGCF 959
Db 896 TEAIENFRTVSLTREQKFEYMAQSLQVPYRNSLRKAHIFGVFSFSTIQAMMYFSYAGCF 955
Qy 960 REGAYLVANEFMNFODVLLVFSIAIVFGAMAVGOVSSFPADYAKAKVSAHVIMIEKSPL 1019
Db 956 REGAYLVAAQLMTFENVMLVFSIAIVFGAMAGNTSSFPADYAKAKVSAHVIMIEKSPL 1015
Qy 1020 IDSYSPHGLKPNLTGNTFNENFVNYTPRDPVPLQGLSLEVKKGQTLALVSGSGCGKS 1079
Db 1016 IDSYSTEGKLPILLGNTFNENFVNYTPRDPVPLQGLSLEVKKGQTLALVSGSGCGKS 1075
Qy 1080 TVVQLLEFYDPLAGSVLIDGKEIKHLNVQWMLRAHLGVISQEPILFDCSIAENIAYGNS 1139
Db 1076 TVVQLLEFYDPMAGSVFLDQKEIKQLNVQWMLRAHLGVISQEPILFDCSIAENIAYGNS 1135
Qy 1140 RVVSHEEIVQAAKEANIHFFIETLPEKYNTRVGDGTQSGGQKORIAIALALVRQPHIL 1199
Db 1136 RAVSHEEIVRAAKEANIHQFIDSLPDKYNTRVGDGTQSGGQKORIAIALALVRQPHIL 1195
Qy 1200 LDEATSLDTSSEKVVQVALDKARKGRITIVIAHRLSTIQNADLIIVVFGQKVEHGT 1259
Db 1196 LDEATSLDTSSEKVVQVALDKARKGRITIVIAHRLSTIQNADLIIVVFGQKVEHGT 1255
Qy 1260 QQLLAOKGIFYSMVSVQAGAKR 1281
Db 1260 QQLLAOKGIFYSMVSVQAGAKR 1281

QY 901 EAIENFRTVSLTROKPEYMYAQLVYPYRNSLRKAHIFGVSPSTQAMMYFSYACGFR 960
|||||
Db 897 EAIENFRTVSLTROKPEYMYAQLVYPYRNSLRKAHIFGVSPSTQAMMYFSYACGFR 956
QY 961 FGALVANEFMFQDVLVFSVAIVFGAMAVGVSSFAFDYAKAKVSAHVIMTIEKSPLI 1020
Db 957 FGALVAVHOIMTFENMLVFSVAIVFGAMAVGVSSFAFDYAKAKVSAHVIMTIEKSPLI 1016
QY 1021 DSYSPRGLKPNLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVGVSSCGGKST 1080
Db 1017 DSYSTRGLKPNLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVGVSSCGGKST 1076
QY 1081 VQVLLERFDYDPLAGSLVDGKEIKHLNVQWLRHAHGLVISOEPILPDCSIAENIAGDMSR 1140
Db 1077 VQVLLERFDYDPLAGSLVDGKEIKHLNVQWLRHAHGLVISOEPILPDCSIAENIAGDMSR 1136
QY 1141 VYSHEIIVAAKEANTHHPETLPKYNTRVGDGTQLSGGQKQRIATARALVROPHILL 1200
Db 1137 VYSQDEIERAANEANHQIESIPDKYNTRVGDGTQLSGGQKQRIATARALVROPHILL 1196
QY 1201 LDATSALDTESEKVVQVQALDKAREGTCIVIAHRLSTIONADLIIVFQNGKVEHGTQ 1260
Db 1197 LDATSALDTESEKVVQVQALDKAREGTCIVIAHRLSTIONADLIIVFQNGKVEHGTQ 1256
QY 1261 OLLAOKGIYFSMVYQAGAKR 1281
Db 1257 OLLAOKGIYFSMVYQAGAKR 1275

RESULT 6
MDRL_RAT STANDARD; PRT; 1277 AA.
AC F43245;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein 1).
GN ABCB1 OR PGV1 OR MDR1 OR MDR1B.
OS Rattus norvegicus (Rat)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92039081; PubMed=1682220;
RA Silverman J.A., Raunio H., Gant T.W., Thorgeirsson S.S.;
RT Cloning and characterization of a member of the rat multidrug
resistance (mdr) gene family.;
RL Gene 106:229-236(1991).
CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- PTM: SEVERAL PHOSPHORYLATED SERINE RESIDUES ARE PRESENT IN THE
LINKER DOMAIN.
CC -1- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
RELATED BUT DISTINCT CELLULAR GENES.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC this SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.
CC
CC -----
CC EMBL; M81855; -. NOT_ANNOTATED_CDS.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmnm.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.

SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 48 71 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 327 347 POTENTIAL.
FT TRANSMEM 348 709 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 710 730 POTENTIAL.
FT TRANSMEM 755 775 POTENTIAL.
FT TRANSMEM 832 852 POTENTIAL.
FT TRANSMEM 854 874 POTENTIAL.
FT TRANSMEM 937 957 POTENTIAL.
FT TRANSMEM 968 988 POTENTIAL.
FT DOMAIN 989 1277 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (BY SIMILARITY).
FT NP_BIND 1070 1077 ATP (POTENTIAL).
FT REPEAT 1 635
FT REPEAT 636 1277
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1277 AA; 141386 MW; 8AFDDDD619D2934C1 CRC64;
Query Match 80.8%; Score 5231.5; DB 1; Length 1277;
Best Local Similarity 80.0%; Pred. No. 3.7e-291;
Matches 1029; Conservative 131; Mismatches 111; Indels 15; Gaps 9;
QY 1 MDPEGRKGSSEKFNWFKMGKSKKKKKKPTVSTFAMPRYSNWLDRVLMVGTMAAI 60
Db 1 MEFEELNGRADKNFSKMGKSKK -EKEKPAVIGTFMFRYADWLKLCMALGTLAAII 58
QY 61 HGAALPLMLVFGNMNDSFANAGISRN-KTPPVIIINESITNTOHFN--HLEEMTYA 117
Db 59 HGTLLPLMLVFGNMTDSTFT--SRPHSDRAITNQSEINST-HTVSDTSLEEDMAMA 114
QY 118 YYYGIGAGVILVAAYIQVSFWCLAGAGQILKIRQFHAIMROEIGWFDVHVGELNRL 177
Db 115 YYYTGIGAGVILVAAYIQVSLWCLAGAGQIHKIRQFHAIMNQEIGWFDVNDAGELNRL 174
QY 178 TDVSKINEGIGDKTGMFFQSIATFTFTGIVGTGKWLTLVILAIAPVLGLSAAIAKI 237
Db 175 TDVSKINDGDKLGMFFQSIATFTFTGIVGTGKWLTLVILAVSPVGLISAMWAKV 234
QY 238 LSSFTDKELLAYAKAGAAVEVLAIRTVIAPFGGKKELRYNNLEAKGIGIKKAITA 297
Db 235 LTFSTNKLQAYAKAGAAVEVLAIRTVIAPFGGKKELRYNNLEAKRVGKIKKAITA 294
QY 298 NISGAAPLLIYASVALAFWGTSLVLSSEVTIGOVLVFFSVLIGAFSIGQASPSIEAF 357
Db 295 NISIGIALLVYASVALAFWGTSLVLSNEYSIGQVLVFFSILLGTFTSIGHLAPNIEAF 354
QY 358 ANARGAAVEIFKIIDNKPSIDSYSGSKGHKPNKGNLEKFNHVSYPKSEVKILKGLNL 417
Db 355 ANARGAAVEIFKIIDNEPSIDSFSTKGHKPDSIMGNLEKFNHVSYPKSEVKILKGLNL 414
QY 418 KYVSGQTVALVNSGCGKSTTVQLMORLDPDIDGMVCDIGQDITINVRHLREITGVVSO 477
Db 415 KYVSGQTVALVNSGCGKSTTVQLLQRLYDPIEGEVSIDGQDITINVRHLREITGVVSO 474
QY 478 EPIVLFATTAENIRYGRNVTWDETEKAVKANAYDFIMKLPNKFTDITLVGERGAOLSGGQ 537
Db 475 EPIVLFATTAENIRYGRNVTWDETEKAVKANAYDFIMKLPNKFTDITLVGERGAOLSGGQ 534
QY 538 KORAIARALVRNPKILLDEATSAIDTSEAVQVVALDKARKGRTTIVIAHRLSTVRNA 597
Db 535 KORAIARALVRNPKILLDEATSAIDTSEAVQVVALDKARKGRTTIVIAHRLSTVRNA 594
QY 598 DVIAGDGDGVIYKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGSKSESDALEMSP 657
Db 595 DVIAGDGGVIVBQGNHEELMKEKGIYFKLVMTQTRGNEIEPCNNAYESQSDTGASELTS 654

```
QY 658 KDSGSLIKRSTRSIIHAPQODRKLGTKEIDLNNVPPVSWFRILKLNSTENPYFVVG 717
Db 655 EESKSPIL-RRSIRSRSHRRQDQRRSSKEDVEDVPMVSFWQILKLNISEWPLYLVG 713
QY 718 FCAINGLOPAFAISFRIIGIFTRDDETPETKRONSMFSLFLVLGIISFTTFPLOGF 777
Db 714 LCAVINGCIQVFAIVSKVGVFSRDDHETKORCNLFSLFLVGMISFVTFYFQGF 773
QY 778 TFGAGEILKRLRYMFRSMKRDVSWFDDPKNTGALTTRLANDAAQVKAIGSLRAY 837
Db 774 TFGAGEILKRLRYMFKMLRDISWFDHKNKTGSLTRLASDASNVKMGMSRLAV 833
QY 838 ITONIANLGTGIIIS--LIYGWOLTLALLAIPIAIAVGVENKMLSGQALKDKELEGA 895
Db 834 VTQNVANLGTGIIISLVYVWGQTLVVIIPIIVLGVGIEMKLLSGQALKDKELETS 893
QY 896 GKIAEAIENFRVTSLTREKFEYMYAQSLQVYPYRNSLRKAHIFGVFSFISITQAMMYFSY 955
Db 894 GKIAEAIENFRVTSLTREKFTETMYAQSLQIPIYRNALKKAHVGTITFAFTQAMMYFSY 953
QY 956 AGCFRGAYLVANEFNFOVLLVFSNAIVFGAMAVQVSSFPADYAKAKVSAHVIMIE 1015
Db 954 AACFRGAYLVARELMTFENVMVLFSAVFGAMAAGNTSSFADYAKAKVSAHSIIIGIE 1013
QY 1016 KSLIDSYSPLGPKNTLEGNVTFNENWYPTRPDIPVLOGLSLEVKKGTALVGS 1075
Db 1014 KIPEIDSYSPLGPKNLEGNVTFNENWYPTRPDIPVLOGLSLEVKKGTALVGS 1073
QY 1076 CGKSTVQVLLERYDPLAGSLVDGKEIKHLNVQMLRAHLGVISQEPILFDCSIAENIAY 1135
Db 1074 CGKSTVQVLLERYDPLAGSLVDGKEIKHLNVQVRA--LGIVSQEPILFDCSIAENIAY 1132
QY 1136 GDSRVVSHIEIVQAAKEANIHHFETLPEKYNTRVGDGTOLSGQKQRTAIARALVRQ 1195
Db 1133 GDSRVVSHIEIVRAAREANIHHFIDSLPEKYNTRVGDGTOLSGQKQRTAIARALVRQ 1192
QY 1196 PHILLDEATSDALTESEKVVQEAIDKAREGTCIVIAHRLSTIONADLVVVFQNGVKVE 1255
Db 1193 PHILLDEATSDALTESEKVVQEAIDKAREGTCVIAHRLSTIONADLVVVFQNGVKVE 1252
QY 1256 HGTHQOLLAQKGIYFSWVSQAGAKR 1281
Db 1253 HGTHQOLLAQKGIYFSW--VQAGAKR 1276

RESULT 7
MDR3 HUMAN
ID MDR3 HUMAN STANDARD; PRT; 1279 AA.
AC P21439;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (p-glycoprotein 3).
GN ABCB4 OR PGY3 OR MDR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89138016; PubMed=2906314;
RA van der Bliek A.M., Kooiman P.M., Schneider C., Borst P.;
RT "Sequence of mdr3 cDNA encoding a human P-glycoprotein.";
RL Gene 71:401-411(1988).
RN [2]
RP SEQUENCE OF 856-1279 FROM N.A.
RX MEDLINE=8911519; PubMed=2892668;
RA van der Bliek A.M., Baas F., ten Houte de Lange T., Kooiman P.M.,
RA van der Velde-Koerts T., Borst P.;
RT "The human mdr3 gene encodes a novel P-glycoprotein homologue and
RT gives rise to alternatively spliced mRNAs in liver.";
RL EMBO J. 6:3325-3331(1987).
```

```
RN GENE STRUCTURE.
RP MEDLINE=91161629; PubMed=2002063;
RA Lincke C.-R. Smit J.J.M., van der Velde-Koerts T., Borst P.;
RT "Structure of the human MDR3 gene and physical mapping of the human
MDR locus.";
RL J. Biol. Chem. 266:5303-5310(1991).
CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS. HUMAN MDR3 IS NOT
CAPABLE OF CONFERRING DRUG RESISTANCE. MEDIATES THE TRANSLLOCATION
OF PHOSPHATIDYLCHOLINE ACROSS THE CANALICULAR MEMBRANE OF THE
HEPATOCTYTE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- DISEASE: DEFECTS IN ABCB4 ARE A CAUSE OF PROGRESSIVE FAMILIAL
INTRAHEPATIC CHOLESTASIS TYPE III (PFIC), A FORM OF AUTOSOMAL
RECESSIVE LIVER DISORDERS, CHARACTERIZED BY EARLY ONSET OF
CHOLESTASIS THAT PROGRESSES TO CIRRHOSIS AND LIVER FAILURE BEFORE
ADULTHOOD.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M23234; AAA36207.1; -.
DR EMBL; X06181; CAA29547.1; -.
DR PIR; JS0051; DVHU3.
DR PIR; A42213; A42213.
DR HSSP; P13569; LNBD.
DR MIM; 171060; -.
DR MIM; 602347; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 192 211 POTENTIAL.
FT TRANSMEM 216 235 POTENTIAL.
FT TRANSMEM 301 320 POTENTIAL.
FT TRANSMEM 336 354 POTENTIAL.
FT DOMAIN 355 711 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 712 732 POTENTIAL.
FT TRANSMEM 756 776 POTENTIAL.
FT TRANSMEM 832 851 POTENTIAL.
FT TRANSMEM 854 873 POTENTIAL.
FT TRANSMEM 937 956 POTENTIAL.
FT TRANSMEM 976 993 POTENTIAL.
FT DOMAIN 994 1279 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 429 436 ATP (BY SIMILARITY).
FT NP_BIND 1069 1076 ATP (BY SIMILARITY).
FT REPEAT 1 640
FT REPEAT 641 1279
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1093 1093 V -> VEVDFGFO (IN REF. 2).
SQ SEQUENCE 1279 AA; 140682 MW; 3D58C98B5C8D6087 CRC64;
```

Query Match 76.0%; Score 4922.5; DB 1; Length 1279;
Best Local Similarity 75.9%; Pred. No. 1.7e-273;
Matches 977; Conservative 129; Mismatches 163; Indels 19; Gaps 9;

QY 1 MDPEGRKG-----SAENFWKMGKKKKKKKKKK-PTVSTFAMPRYSNWLDRLYMLV 53
Db 1 MDLEAKNGTAWRPTSAGDF-ELGSSQKQRKTKTKVMGLVTLFYSQWQDKLPSL 59
QY 54 GTMAAIHGAALPLMLVFGNMTDSFANAGISRNKTPFVIINESITNNTQHFINHLEEM 113
Db 60 GT1MAIHGSGPLPMWIVGEMTKDFVT--AGNFSFPVNFSLSLNPGK----ILBEE 113
QY 114 TTYAYYSIGAGVLVAAYIQVSFCLAGROILKIRQFPHAIMRQIGLWFDVHVGEL 173
Db 114 TRYAYYSGLGAGVLVAAYIQVSFTLAAGROIRKIRQFFHAILRQIGLWFDITTEL 173
QY 174 NTRLTDDVSKINEGTGDKTGNFQSIATFETGFIYVGFTRGKMLTLVILAISPVLGLSAAI 233
Db 174 NTRLTDDISKISIEGIGDKRGVGFQAVATFFAGFIYVGFIRGKMLTLVIMAILSPILGSAV 233
QY 234 WAKILSSFTDKELLAYAKAGAAVEVLAARTVIAFGQKKKELERYKNLEAKGIGTKK 293
Db 234 WAKILSAFSDKELAYAKAGAAVEALGAIRTVIAFGQNKELERYQKHLNAKEIGIKK 293
QY 294 ATANISIGAAPLLIYASALAFWGTSLVLSSEYTGIVLTVFTSVLIGAFSIGQASPS 353
Db 294 AITSANSMGIAFLLIYASALAFWGSTLVISKEYTIGNAMTVFVSILIGAFSVGQAAPC 353
QY 354 IEAFANARAAVEIFRIIDONKPSIDSYSKSHGHPDNKGNHFSYPSKEVKILK 413
Db 354 IDAFANARAAVIFDIIDONKPIDSFSGHGPDSIKGNLEFNDVHFSYPSRANVILK 413
QY 414 GLNLKVSQGTVALVNSGCGSTTVOLMORLYDPTDGMVCIDGQDRTINVRHLEITG 473
Db 414 GLNLKVSQGTVALVNSGCGSTTVOLLRQYLDPEGTINIDGQDIRNFNYLREIIG 473
QY 474 VVSQEPVLFATTIENIRYGRNVTWDETEKAVKEANAYDFIMKLPNKFTDLVGERGAOL 533
Db 474 VVSQEPVLFSTTIAENICIRYGRNVTWDETEKAVKEANAYEFIMKLPQKEDTLVGERGAOL 533
QY 534 SGGQKQRIATARALVRNPKILLDEATSAIDTESEAVVQVLDKARKGRTTIVIAHRLST 593
Db 534 SGGQKQRIATARALVRNPKILLDEATSAIDTESEAEVQAALDKAREGRTTIVIAHRLST 593
QY 594 VRNADVITAGDDGVIVKEGNHDELMKEGYIFKLVTWQTRGNEIELEENATGESKESDAL 653
Db 594 VRNADVITAGDDGVIVGEGSHLSMKEGYIFKLVTWQTRGNEIELEENATGESKESDAL 651
QY 654 EMPKDSGSLIKRRSTRIRSIHAPOGDRKLGTKED--LAVENVPVFWRLKLNSTEWPY 712
Db 652 RMAPNGWKSRLF-RHSTQKNLKSQMCOKSLDVTGDLNENVPVFWRLKLNSTEWPY 710
QY 713 FVVGIFCAIINGLQPAFSIFRSIIGFTFREDPETKRONSMNFSVLFLVIGLISFTIF 772
Db 711 FVVGTVCAIANGLQPAFSVIFSEIIAIFGPGDD-AVKQKCNIFSLFLFLGLIISFFTF 769
QY 773 FLOGFTGKAGEILTKRLRYWFRSMRLQDVSNFDDPKNTTGALTURLANDAQAQVKAIG 832
Db 770 FLOGFTGKAGEILTKRLRYWFRSMRLQDVSNFDDPKNTTGALTURLANDAQAQVKAIG 829
QY 833 SRLAVITQNIANTLGTIIISLIYQWQLTLLLAIVPIIATAGVEMKMSGQALKDKKEL 892
Db 830 TRLALIAQNIANTLGTIIISFIYQWQLTLLLAIVPIIATAGVEMKMSGQALKDKKEL 889
QY 893 EGAKKIATEAIENRTVSLTREQKFEYMAQSLQVYRNLSRKAHFGVFSFISITQAMMY 952
Db 890 EAAGKIATEAIENRTVSLTREQKFEYMAQSLQVYRNLSRKAHFGVFSFISITQAMMY 949
QY 953 FSYAGCFRFGAYLVANEFMFQDVLVFSALVFGAMAVGVSSFAPDYAKAKYSAAHVIM 1012
Db 950 FSYAGCFRFGAYLVANEFMFQDVLVFSALVFGAMAVGVSSFAPDYAKAKYSAAHVIM 1009
QY 1013 ITIEKPSIDSYSPHGLKPNLTLEGWNTNEVVNTPTRPDLPVLOGLSLEKVKGTALV 1072
Db 1010 LFERQPLIDSYSSEGLKPKDEGNITTFNEVVNTPTRPDLPVLOGLSLEKVKGTALV 1069
QY 1073 SSGCGKSTVVQLLRFYDPLAGSVLIDGKEIKHLNVQWLAHGLVIGVSEPIFLDCSIAEN 1132

Db 1070 SSGCGKSTVVQLLRFYDPLAGSVLIDGKEIKHLNVQWLAHGLVIGVSEPIFLDCSIAEN 1129
QY 1133 IAYGDNRSRVVSHIEVOAAEKANHFIETLPEKYNTRYDGTQSLSGGQKQRIATARAL 1192
Db 1130 IAYGDNRSRVVSHIEVOAAEKANHFIETLPEKYNTRYDGTQSLSGGQKQRIATARAL 1189
QY 1193 VROPHILLDEATSAIDTESEKVVQALDKAREGRTCTIVIAHRLSTIONADLLVVPQNK 1252
Db 1190 IROPQILLDEATSAIDTESEKVVQALDKAREGRTCTIVIAHRLSTIONADLLVVPQNR 1249
QY 1253 VKEHGTHQOQLLAOKGIYFSVMVQAGAK 1280
Db 1250 VKEHGTHQOQLLAOKGIYFSVMVQAGTQ 1277
RESULT 8
MDR2_MOUSE STANDARD; PRT; 1276 AA.
ID MDR2_MOUSE AC P21440;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 2 (p-glycoprotein 2).
GN PGY2 OR PGY-2 OR MDR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86302195; PubMed=3405218;
RT Gros P., Raymond M., Bell J., Housman D.;
RT "Cloning and characterization of a second member of the mouse mdr
RT gene family.";
RL Mol. Cell. Biol. 8:2770-2778(1988).
RN [2]
RP SEQUENCE OF 1-23 FROM N.A.
RC STRAIN=BALB/C;
RA Kirschner L.S., Horwitz S.B.;
RL Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS. MOUSE MDR2 IS
CC NOT CAPABLE OF CONFERRING DRUG RESISTANCE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
CC RELATED BUT DISTINCT CELLULAR GENES.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J03398; AAA39516.1; -
CC EMBL: M7451; AAA39515.1; -
CC PIR: A30409; DVMS2.
CC HSSP: P13569; INBD.
CC MGD: MGI:97569; Pgy2.
CC InterPro: IPR001140; ABC_transporter_tmemb.
CC InterPro: IPR003439; ABC_transporter.
CC InterPro: IPR001687; ATP_GTP_A.
CC Pfam: PF00664; ABC_membrane_2.
CC Pfam: PF00005; ABC_tran; 2.
CC SMART: SM00382; AAA; 2.
CC PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 52 75 POTENTIAL.
 FT TRANSMEM 119 139 POTENTIAL.
 FT TRANSMEM 188 208 POTENTIAL.
 FT TRANSMEM 215 235 POTENTIAL.
 FT TRANSMEM 296 316 POTENTIAL.
 FT TRANSMEM 325 345 POTENTIAL.
 FT TRANSMEM 346 707 POTENTIAL.
 FT TRANSMEM 708 728 POTENTIAL.
 FT TRANSMEM 753 773 POTENTIAL.
 FT TRANSMEM 829 849 POTENTIAL.
 FT TRANSMEM 850 870 POTENTIAL.
 FT TRANSMEM 933 953 POTENTIAL.
 FT TRANSMEM 970 990 POTENTIAL.
 FT TRANSMEM 991 1276 POTENTIAL.
 FT NP_BIND 426 433 ATP (BY SIMILARITY).
 FT NP_BIND 1066 1073 ATP (BY SIMILARITY).
 FT REPEAT 1 635
 FT REPEAT 636 1276
 FT REPEAT 1276 AA; A6C38DCD2C118EEF CRC64;
 SQ SEQUENCE 1276 AA; 140332 MW; A6C38DCD2C118EEF CRC64;
 Query Match 75.2%; Score 4870; DB 1; Length 1276;
 Best Local Similarity 74.9%; Pred. No. 1.7e-270;
 Matches 962; Conservative 138; Mismatches 170; Indels 14; Gaps 8;
 QY 1 MDPEGRKGSNAE--NFWKMGKSKK-KEKKEKPTVSTFAMFRYSNWLDRLYMLVGTMA 57
 DB 1 MDLEAARNGTARRLDGDFELGSIQNOGREKKKVNLLGLTLFRYSQWQDKLFLMELGTLM 60
 QY 58 AIHGAALPLMLVFGNMTDSEA-NAGISRNKTFPPVLINESTINNTQHFHNHLEEMTTY 116
 DB 61 AIHAGSLPLMLVFGNMTDFTDNTG---NFSLPNFSLSMLNPGR---ILEEEMTRY 113
 QY 117 AYYSGIGAGVLVAAYIQVSWFCLAGROILKIRKOPFFHAIMROEIGWFDVHDVGLNTR 176
 DB 114 AYYSGIGAGVLVAAYIQVSWFCLAGROILKIRKOPFFHAIMROEIGWFDVHDVGLNTR 173
 QY 177 LTDDVSKINGIGDKIMFQSTATFTFTGIVGTGRGKTLVILALSPVLGSLAAIWA 236
 DB 174 LTDDVSKINGIGDKIMFQSTATFTFTGIVGTGRGKTLVILALSPVLGSLAAIWA 233
 QY 237 ILSFTDKELAYAKAGAAVEEVLAAIRTVIAFGGKKELERYKNLEAKGIGIKKAIT 296
 DB 234 ILSFTDKELAYAKAGAAVEEVLAAIRTVIAFGGKKELERYKNLEAKGIGIKKAIT 293
 QY 297 ANTISGAFLLIYASYALAFWYGTSLVSEYITIGOVTVFFSVLIGAFSIGOASPSIA 356
 DB 294 ANTISGAFLLIYASYALAFWYGTSLVSEYITIGOVTVFFSVLIGAFSIGOASPSIA 353
 QY 357 FANARGAAYEIFKIIDNKPSIDSYKSGHKPDNIKONLEKFNHVSYPKREKYLKGLN 416
 DB 354 FANARGAAYEIFKIIDNKPSIDSYKSGHKPDNIKONLEKFNHVSYPKREKYLKGLN 413
 QY 417 LKVSQGTVALVNSGCGSTTVOLMORLYDPTDGMVCDIGDITINVRHLREITGVYS 476
 DB 414 LKVSQGTVALVNSGCGSTTVOLMORLYDPTDGMVCDIGDITINVRHLREITGVYS 473
 QY 477 QEPVLPAITAEINIRYGRNVTMDTEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSG 536
 DB 474 QEPVLPAITAEINIRYGRNVTMDTEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSG 533
 QY 537 OKORIAIARALVRNPKILLDEATSDLTSEAVVQVALDKARKGRTTIVIAHRLSTVRN 596
 DB 534 OKORIAIARALVRNPKILLDEATSDLTSEAVVQVALDKARKGRTTIVIAHRLSTVRN 593
 QY 597 ADVIAGFDGIVVEKGNHDELKKEGIYFKLVTMTGRNEIELEATGESKESDALEMS 656
 DB 594 ADVIAGFDGIVVEKGNHDELKKEGIYFKLVTMTGRNEIELEATGESKESDALEMS 653
 QY 657 PKDSSGLLKRRSTRSHAPQDQKLGTKEDLNENPPVPSFWRLKLNSTEWFFVVG 716
 DB 654 P-NGWKARIFNRSTKSLKSPH-QNRLDEETNELDANPPVPSFLVKLNKTEWFFVVG 711
 QY 717 IFCAIINGLOPAFSIIIRIIGITRDEDDPTKQNSNMFSVLFLVLGIISFITFFLQ 776

Db 712 TVCAIANGALQPAFSIIILSEMIAIFGPGDD-AVKQKCNMFSVLFLGLVLSFTFFLQ 770
 QY 777 FTFGKAGEILTKRLRYMVFRLQVSWFDDPKNTTGTALTTLANDAAQVKAIGSRLA 836
 Db 771 FTFGKAGEILTKRLSRMAFKAMLRQDSWFDHDKNSTGALSTRLAIDAQVQATGKLA 830
 QY 837 VIQTANTLGTGIIISIIYQWQLTLLLLAIVPIIAIAGVVMKMLSGQALKDKKELEG 896
 Db 831 LIAQNTANLGTGIIISIIYQWQLTLLLSVVPPIAVAGIVEMKLAGNAKRDKEAAG 890
 QY 897 KIATEAENFTVSLTREQFMYAQSLQVPRYNSLRKHAHFGVGSFISITQAMVFSYA 956
 Db 891 KIATEAENIRTVSTQTKESMIVKELHGPYRNSVRKHAHIGITFSIQAFVFSYA 950
 QY 957 GCFRGAYLVANEFMNFQDVLVFSIAIVFGAMAVGVSSFAPDYAKAKVSAARHVMIEK 1016
 Db 951 GCFRGAYLVANEFMNFQDVLVFSIAIVFGAMAVGVSSFAPDYAKAKVSAARHVMIEK 1010
 QY 1017 SPLIDSYSRGLKPNVLEGNVTFNENFVNYPRDIPVLOGLSLEVKKGTTLALVGSSG 1076
 Db 1011 QPLIDSYSRGLKPNVLEGNVTFNENFVNYPRDIPVLOGLSLEVKKGTTLALVGSSG 1070
 QY 1077 GKSTVQVLLERFYDPLAGSVLIDGKEIKHLNVQWLRHGLIVSOEPIPLFDCSIAENIAG 1136
 Db 1071 GKSTVQVLLERFYDPLAGSVLIDGKEIKHLNVQWLRHGLIVSOEPIPLFDCSIAENIAG 1130
 QY 1137 DNSRVVSHBEIYVQAAKEANHHFIETLPKYNTRYGDKGTQSLSGGKORIALARALVRP 1196
 Db 1131 DNSRVVSHBEIYVQAAKEANHHFIETLPKYNTRYGDKGTQSLSGGKORIALARALVRP 1190
 QY 1197 HILLDEATSDLTSEKVVQVQALDKAREGRTCVIAHRLSTIQNALDVIIVFONGKVEH 1256
 Db 1191 HILLDEATSDLTSEKVVQVQALDKAREGRTCVIAHRLSTIQNALDVIIVFONGKVEH 1250
 QY 1257 GTHQQLLAQGIYFNSVMSVQAGAK 1280
 Db 1251 GTHQQLLAQGIYFNSVMSVQAGAK 1274
 RESULT 9
 MDR3_CRIGR
 ID MDR3_CRIGR STANDARD; PRT; 1281 AA.
 AC P23174;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Multidrug resistance protein 3 (P-glycoprotein 3).
 GN PG3 OR PG3.
 OS Cricetulus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92135896; PubMed=1685679;
 RA Endicott J.A., Sarangi F., Ling V.;
 RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
 gene family.";
 RL DNA Seq. 2:89-101(1991).
 CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
 CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT
 CC CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
 CC CANNOT.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M60042; AAA68885.1; -
DR HSP; P13569; 1NBD.
DR InterPro; IPR001140; AAA.
DR InterPro; IPR001140; ABC_transporter_tmcm.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 299 319 POTENTIAL.
FT TRANSMEM 328 348 POTENTIAL.
FT DOMAIN 349 712 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 713 733 POTENTIAL.
FT TRANSMEM 758 778 POTENTIAL.
FT TRANSMEM 834 854 POTENTIAL.
FT TRANSMEM 855 875 POTENTIAL.
FT TRANSMEM 938 958 POTENTIAL.
FT TRANSMEM 975 995 POTENTIAL.
FT DOMAIN 996 1281 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 429 436 ATP (POTENTIAL).
FT NP_BIND 1071 1078 ATP (POTENTIAL).
SQ SSEQUENCE 1281 AA; 140866 MW; 2203EF61EBB29602 CRC64;

Query Match 74.6%; Score 4829.5; DB 1; Length 1281;
Best Local Similarity 73.8%; Pred. No. 3.4e-268;
Matches 951; Conservative 146; Mismatches 174; Indels 17; Gaps 8;
QY 1 MDPEGRKGSAAK-----NFWKMGKSKK-KEKKEKPTVSTAMFRRYSNWLDRMLVYG 54
DB 1 MDLEAARNGTARRPGTVGDFSGISNQGRRKKKVNKLIGLPTLPRYSQWQKLMUG 60
QY 55 TMAAIIHGAALPLMLVFGNMTDSFA-NAGISRNKTPFPVILINESITNTQHFHINLEEM 113
DB 61 TMAIAHSGSLPLMMIVFGEMTDKFNAG--NFSLPVNFSLSMINPGR----ILEEM 113
QY 114 TTYAYYSGAGVLVAAYIQVFWCLAAQRQILKIRKQFFHAIMQEIQWDFVDHVGEL 173
DB 114 TRYAYYSGLGGVLAAYIQVFWTLAAGRQIKKIRQNFHAILRQEMGWDFDKGTTEL 173
QY 174 NTRLTDDVSKINEGIGDKIGMFTQSIAFTFTGTVGTRGKLTVLTAISPVLGLSAAI 233
DB 174 NTRLTDDISKISEGIGDKVGFQVATVFFAGTVGFIQWKLTLVTAISPILGLSAAV 233
QY 234 WAKILSFTDKELAYAKAGAAVEELAAIRTVIAFGQKKELERYNNKLEAKGIGIKK 293
DB 234 WAKILSFTDKELAYAKAGAAVEELAAIRTVIAFGQKKELERYNNKLEAKGIGIKK 293
QY 294 AITANISGAFLIYASALAFWYGSYLVLSSEYTTIGOVLTFFSVLIGAFSIGQASPS 353
DB 294 AISANISGIAFLIYASALAFWYGSYLVLSSEYTTIGOVLTFFSVLIGAFSIGQASPS 353
QY 354 IEAFANARGAAEYFKIIDNKPDSYKSGKPDNKGKLEKFNKRVFSPKRVKILK 413
DB 354 IDAFANARGAAEYFKIIDNKPDSYKSGKPDNKGKLEKFNKRVFSPKRVKILK 413
QY 414 GLNLKVSQGVVALVNGSGCKSTTVQLMORLXPDTCGMVICIDQDRTINVRHLRITG 473
DB 414 GLNLKVSQGVVALVNGSGCKSTTVQLMORLXPDTCGMVICIDQDRTINVRHLRITG 473
QY 474 VVSOEPLVFAFTIENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFOTLVGERGAQL 533
DB 474 VVSOEPLVFAFTIENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFOTLVGERGAQL 533

DB 474 VVSOEPLVFAFTIENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFOTLVGERGAQL 533
QY 534 SGGQKQRIATARALVRNPKILLDEATSDLTSEAEVVOVALDKARKGRITIVIAHRLST 593
DB 534 SGGQKQRIATARALVRNPKILLDEATSDLTSEAEVVOVALDKARKGRITIVIAHRLST 593
QY 594 VRNADVIAGFDGIVVEKGNHDELMEKGIYFKLVMTQTRGNEIELEENATGESKSDAL 653
DB 594 VRNADVIAGFDGIVVEKGNHDELMEKGIYFKLVMTQTRGNEIELEENATGESKSDAL 653
QY 654 EMSPKDGSSSLIKRRSTRSIHAPQGDQKRLGTKEK-LNENVPVSPFWRLTKLNSTWMPY 712
DB 654 GMP-NGWKSHIFRNSTKSLKSSRAHHRLDVADELDAVPVFLKLNKLNKTEWMPY 712
QY 713 FVVGICAILNGLOPAFISIFRIIGIFRDEPETKRONSMNMFSLFLVLGLIISFITF 772
DB 713 FVVGICAILNGLOPAFISIFRIIGIFRDEPETKRONSMNMFSLFLVLGLIISFITF 772
QY 773 FLOGFTGKAGEILTTLRLSMFAKMLRQDMWDFDDYKNSGTALSTRLATRAQVQATG 831
DB 773 FLOGFTGKAGEILTTLRLSMFAKMLRQDMWDFDDYKNSGTALSTRLATRAQVQATG 831
QY 833 SRLAVITONIANLGTIIISLYGWLTLALLAIVPIIAGVVMKLSQALKDKKEL 892
DB 833 TRALIAQNTANLGTIIISLYGWLTLALLAIVPIIAGVVMKLSQALKDKKEL 892
QY 893 EGAGKIATEAIENFRTVVSILTREQFEYMAQSLQVYPYRNSLRKAHIFGVSFSTQAMMY 952
DB 893 EGAGKIATEAIENFRTVVSILTREQFEYMAQSLQVYPYRNSLRKAHIFGVSFSTQAMMY 952
QY 953 PSYACGFRGAYLVANEFMNFQDVLVPSAIVFGMAVGQVSSFAPDYAKAKVSAAHVIM 1012
DB 953 PSYACGFRGAYLVANEFMNFQDVLVPSAIVFGMAVGQVSSFAPDYAKAKVSAAHVIM 1012
QY 1013 IIEKSPILDSYSPHGLKPTLEGVNTFNEVFNTPTRDIPVLQGLSLEVKKGTALVG 1072
DB 1013 IIEKSPILDSYSPHGLKPTLEGVNTFNEVFNTPTRDIPVLQGLSLEVKKGTALVG 1072
QY 1073 SSGCKSTVQVQLLERYDPLAGSLVDIGKEIKHLNVQMLRAHLGVSEPTLFDPCSTAE 1132
DB 1073 SSGCKSTVQVQLLERYDPLAGSLVDIGKEIKHLNVQMLRAHLGVSEPTLFDPCSTAE 1132
QY 1133 IAYGNSRVVSHETVQAAKANIHHFTETLPEKYNTRVGDGKQTLGGQKQRIARAL 1192
DB 1133 IAYGNSRVVSHETVQAAKANIHHFTETLPEKYNTRVGDGKQTLGGQKQRIARAL 1192
QY 1193 VRQPHILLDEATSDLTSEKVKQALDKAREGRTCTIVIAHRLSTIQNADLIIVVFQNGK 1252
DB 1193 VRQPHILLDEATSDLTSEKVKQALDKAREGRTCTIVIAHRLSTIQNADLIIVVFQNGK 1252
QY 1253 VKEHGHQOALLAOKGIYFSWVSQAGAK 1280
DB 1253 VKEHGHQOALLAOKGIYFSWVSQAGAK 1280
RESULT 10
MDR2_RAT
ID MDR2_RAT STANDARD; PRT; 1278 AA.
AC Q08201;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 2 (P-glycoprotein 2).
GN PGY2 OR MDR2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISCHER;
RX MEDLINE=93376516; PubMed=8103593;
RA Brown P.C., Thorgeirsson S.S., Silverman J.A.;

RT "Cloning and regulation of the rat mdr2 gene.";
RL Nucleic Acids Res. 21:3885-3891(1993).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC SUBCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L15079; AAA02937.1; -
DR HSSP; P13569; INBD
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 54 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 55 75 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT DOMAIN 346 709 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 710 730 POTENTIAL.
FT TRANSMEM 755 775 POTENTIAL.
FT TRANSMEM 831 851 POTENTIAL.
FT TRANSMEM 852 872 POTENTIAL.
FT TRANSMEM 935 955 POTENTIAL.
FT TRANSMEM 972 992 POTENTIAL.
FT DOMAIN 993 1278 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (POTENTIAL).
FT NP_BIND 1068 1075 ATP (POTENTIAL).
SQ SEQUENCE 1278 AA; 140655 MW; D4FB6BE745AF73BF CRC64;

Query Match 74.4%; Score 4816; DB 1; Length 1278;
Best Local Similarity 73.9%; Pred. No. 2e-267;
Matches 950; Conservative 145; Mismatches 176; Indels 14; Gaps 8;

QY 1 MDEPGKGSAAEK--NFKMGK-KSKKKEKKEKPTVSTFAMFRYSNNMLDRMLVGMMA 57
DB 1 MDLEAARNGTARRLDGDFELGTSISNGSREKKKVNLI GPLTLFRSDMQDKLFMLLGTA 60

QY 58 AIITHGAALPLMLVFCNWDSTA-NAGSRNKTFFVINESTNTNQHFINHEEEMTY 116
DB AIATHGSLPLMLVFCGMDTKFDVAG---NFSLPVNFSLSLNPNR----ILEEMTRY 113

QY 117 AYYISGIGAGVLAAYIQVSWCLAGROILKIRKOFFHATMRQIGHFWDVHVGELNTR 176
DB AYYISGLGGVLLAAYIQVSWFLLAAGROIRKIRKOFFHAILROBMGFIDKGTTELNR 173

QY 177 LTDVSKINIGDKIGMFPQSIAFTFTGTFVGRGKLTILVILATSPVLGLSAAIWA 236
DB LTDDISKISGIGDKVGMFFQAIATFFAGFIVGFRGKLTILVILATILGLSTAVWAK 233

QY 237 ILSSTFDKELLAYAKAGAAVEVLAAITVIAFGQKKELERYNNKLEAKGIGIKKAIT 296
DB ILSSTFDKELLAYAKAGAAVEALGAIRTVIAFGQKKELERYNNKLEAKGIGIKKAIT 293

QY 297 ANISGIAFLLIYASYALAFWGTSLVLSSEYTGIVLTVFFSLIGAFSGQASPIEA 356
DB ANISGIAFLLIYASYALAFWGTSLVLSSEYTGIVLTVFFSLIGAFSGVQAAPCIDA 353

QY 357 FANARGAAEYFIFKIIDNKPISDSYSKSGHKPDNKGKLEFNHFSYSPSRKEVKILKGLN 416
DB FPNARGAAVYFIDFIDNNPKIDFSEGRHKPDSIKGNLEFSDVHFSYSPSRANIKILKGLN 413

QY 417 LKVGSGQVALVGNSSGCKSTTVOLMORLYDPTDGMVCIDGQDRTINVRHLREITGVVS 476
DB LKVGSGQVALVGNSSGCKSTTVOLMORLYDPTDGMVCIDGQDRTINVRHLREITGVVS 473

QY 477 QEPVLFAITAEINIRYGRNVMTDEIEKAYKEANAYDFIMKLPNKFDTLVGERGAOLSGS 536
DB QEPVLFAITAEINIRYGRNVMTDEIEKAYKEANAYDFIMKLPNKFDTLVGERGAOLSGS 533

QY 537 QKQRIATARALVRNPKILLDEATSDTSEAVVQVALDKARKGRTTIVIAHRLSTVRN 596
DB QKQRIATARALVRNPKILLDEATSDTSEAVVQVALDKARKGRTTIVIAHRLSTVRN 593

QY 597 ADVIAGFDGIVVEKGNHDELMKEGIYFKLVMTQRTGRNEIELENATGESKESDALEMS 656
DB ADVIAGFDGIVVEKGNHDELMKEGIYFKLVMTQRTGRNEIELENATGESKESDALEMS 653

QY 657 PKDSSSLIKRRSTRRSIHAPOQDRLKGTGK-EDLNENVPVSWFRILKLNSTWPPYFV 715
DB P-NGWKARIFRNSYTKSLKSSRAHQNRDLVETNELDANVPVSWFLVRLNKTEWPPYFV 712

QY 716 GIFCAIINGLOPAFASIIISRIIGITFRDDETPETKQNSMPSVFLVLIIGISITFFLQ 775
DB GTLCAIANGALOPAFASIIISRIIGITFRDDETPETKQNSMPSVFLVLIIGISITFFLQ 771

QY 776 GFTFGKAGEILTKRLRYMFRSMRDVDFDDPKNTGALTTLRLANDAAQVKGASRL 835
DB GFTFGKAGEILTKRLRYMFRSMRDVDFDDPKNTGALTTLRLANDAAQVKGASRL 831

QY 836 AVITQNTANLGTGIISLIYWGQTLTLLLAIVPIIAAGVVMKMLSGQALDKKLEGA 895
DB ALIAQNTANLGTGIISLIYWGQTLTLLLAIVPIIAAGVVMKMLSGQALDKKLEGA 891

QY 896 GKATEAIENFRVVSILTREQEYMYAQSLOVPYRNSLRKAHIFGVFSITQAMWYFSY 955
DB GKATEAIENFRVVSILTREQEYMYAQSLOVPYRNSLRKAHIFGVFSITQAMWYFSY 951

QY 956 AGCFRFGAYLVANEFMNFODVLLVFAIVFGAMAVGOVSSFPADYAKAKVSAHVIMLIE 1015
DB AGCFRFGAYLVANEFMNFODVLLVFAIVFGAMAVGOVSSFPADYAKAKVSAHVIMLIE 1011

QY 1016 KSPIDISYSPHGLKPNLTLEGNVTFNEVFNPTPRDIPVQLGLSLEVKKGQTLALVSGSG 1075
DB KSPIDISYSPHGLKPNLTLEGNVTFNEVFNPTPRDIPVQLGLSLEVKKGQTLALVSGSG 1071

QY 1076 CGKSTVVQLLERYDPLAGSVLIDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSTAEINTAY 1135
DB CGKSTVVQLLERYDPLAGSVLIDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSTAEINTAY 1131

QY 1136 GDNRSVVSHEEIVQAAKEANIHHFETLPEKYNTRVKGDTQSLGQKQRIATARALVQ 1195
DB GDNRSVVSHEEIVQAAKEANIHHFETLPEKYNTRVKGDTQSLGQKQRIATARALVQ 1191

QY 1196 PHILLDEATSDTSEKVKQVQALDKAREGRTTIVIAHRLSTIONADLIIVFONGVKYE 1255
DB PHILLDEATSDTSEKVKQVQALDKAREGRTTIVIAHRLSTIONADLIIVFONGVKYE 1251

QY 1256 HGTHQOLLAQKGIYFSWVQAGAK 1280
DB HGTHQOLLAQKGIYFSWVQAGAK 1276

RESULT 11
AB11.HUMAN
ID AB11.HUMAN STANDARD; PRT; 1321 AA.
AC O95342; Q9UNB2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11).
GN ABCB11 OR BSEP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., AND VARIANT PFIC2 GLY-297.
RP MEDLINE=99021377; PubMed=9806540;
RA Strautnieks S.S., Bull L.N., Knisely A.S., Kocoshis S.A., Dahl N.,
RA Arnell H., Sokal E., Dahan K., Childs S., Ling V., Tanner M.S.,
RA Kagalwalla A.F., Nemeth A., Pawlowska J., Baker A., Mieli-Vergani G.,
RA Freider N.B., Gardiner R.M., Thompson R.J.
RT A gene encoding a liver-specific ABC transporter is mutated in
RT progressive familial intrahepatic cholestasis.;
RL Nat. Genet. 20:233-238(1998).
[2]
RN SEQUENCE FROM N.A.
RP Mol O., Hooiveld G.J.E.J., Jansen P.L.M., Muller M.;
RA Cellular localization and functional characterization of the human
RT bile salt export pump (BSEP).;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
CC INTO THE CANALICULUS OF HEPATOCYTES.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
CC MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
CC SITU (BY SIMILARITY).
CC -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -!- DISEASE: DEFECTS IN ABCB11 ARE THE CAUSE OF PROGRESSIVE FAMILIAL
CC INTRAHEPATIC CHOLESTASIS 2 (PFIC2), AN INHERITED LIVER DISEASE OF
CC CHILDHOOD. PFIC2 IS CHARACTERIZED BY CHOLESTASIS AND NORMAL SERUM
CC GAMMA-GUTAMYLTRANSFERASE ACTIVITY.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF091582; AAC77455.1; -;
DR EMBL; AF136523; AAD28285.1; -;
DR MIM; 603201; -;
DR MIM; 601847; -;
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmemb.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transmembrane; Transport; Disease mutation.
FT DOMAIN 1 62
FT TRANSFAM 63 83
FT DOMAIN 84 147
FT TRANSFAM 148 168
FT DOMAIN 169 215
FT TRANSFAM 216 236
FT DOMAIN 237 240
FT TRANSFAM 241 261
FT DOMAIN 262 319
FT TRANSFAM 320 340
FT DOMAIN 341 353
FT TRANSFAM 354 374
FT DOMAIN 375 755
FT TRANSFAM 756 776
FT TRANSFAM 776 776

FT	DOMAIN	777	794	EXTRACELLULAR (POTENTIAL).
FT	TRANSFAM	795	815	POTENTIAL.
FT	DOMAIN	816	869	CYTOPLASMIC (POTENTIAL).
FT	TRANSFAM	870	890	POTENTIAL.
FT	TRANSFAM	891	911	POTENTIAL.
FT	DOMAIN	912	979	CYTOPLASMIC (POTENTIAL).
FT	TRANSFAM	980	1000	POTENTIAL.
FT	DOMAIN	1001	1011	EXTRACELLULAR (POTENTIAL).
FT	TRANSFAM	1012	1032	POTENTIAL.
FT	DOMAIN	1033	1321	CYTOPLASMIC (POTENTIAL).
FT	NP_BIND	455	462	ATP (POTENTIAL).
FT	NP_BIND	1113	1120	ATP (POTENTIAL).
FT	CARBOHYD	109	109	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	122	122	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	297	297	E -> G (IN PFIC2).
FT	CONFLICT	339	339	/FTid=VAR_010271.
FT	CONFLICT	444	444	V -> L (IN REF. 2).
FT	CONFLICT	444	444	V -> A (IN REF. 2).
FT	SEQUENCE	1321	AA; 146392	MW; D444CACC48DEA371 CRC64;

Query Match 51.1%; Score 3308.5; DB 1; Length 1321;
Best Local Similarity 50.8%; Pred. No. 2.9e-181;
Matches 654; Conservative 241; Mismatches 359; Indels 33; Gaps 9;

QY	20	KSK-KKEKKEKPTSTFAMFRYSNMLDRMLVGMVMAIIHGAALPLMLVGNMTDS	78
DB	30	KKSRLQDEKKGVRVGFQFLFRSSSDIWLMEVSGLCAPFLHGAQPGVLLFGTMTDV	89
QY	79	FANAGISR-----NKTFF---PVINESNTNTOHFINHLEEMTYAYVYSGIG	124
DB	90	FIDYDVELOEQIPGKACVNNITVWNSLSNQNMNTGRCGLLNIESEMIKFSYAGIA	149
QY	125	AGLVAAVYQVFWCLAAAGRIKIRKOFFHAIMRQEGWFDVHDVGBELNRLRDDVSKI	184
DB	150	VAVLTGVIQICFWIAAARQIQMRKPYFRIRIMRMEIGWDFCNSVGLNTRFDDINKI	209
QY	185	NEGIGDKTGMFFQSTATFTFTGVTGKWLTVILASPVGLSIAWAKILSFTDK	244
DB	210	NDAIADQWALFQRTMTSTCTGFLGFRGKWLTVIISVPLIGIGATIGLSVSKFTDY	269
QY	245	ELLAYAKAGAAVEALVIAFTVAFGGQKELERYKNKLEAKGIGIKKATANISIGAA	304
DB	270	ELKAYAKAGVADVISMMRTVAASGGEKREVERKKNLVAQRWGIKRGIVMGFFTFV	329
QY	305	FLLIYASVALAFWYGTSLVL-SSEYITQVLTVPFVSLIGAFSIGQASPSIEAFANAGA	363
DB	330	WCLIFLCYAVAFWYGSTLVLDGEYTPGTVLQIFLSVIGALNLGNASPCLEAFATGRA	389
QY	364	AYEIEKIDNKPSIDSYSGSKGHKPNIKGNLFKNVHFSYPSRKEVKILGLNLKVSQSG	423
DB	390	ATSFETIDRKPIIDCMSEDGKLDRIKEIEFHNVTFFPSRPEVKILNDLNMVVKPGE	449
QY	424	TVALVNSCGKSTTVQLMORLYDPTDGMVCIDGDIPTFNVHRHLREITGVVSPQVFLA	483
DB	450	MTALVPGSGAGKSTALQLIQREYDPCGMVTVDGHDIRSLNQWLDRDQIGIVEQEPVLF	509
QY	484	TTIAENIRYGRNVNMDIEKAVKANAYDFIMKLPKPFDTLVGRGAQLSGGQKORIAI	543
DB	510	TTIAENIRYGRDATMEDIVQAAKANAYNFIMDLPQPDFTLVGGGGOMSGGQKORIAI	569
QY	544	ARALVNRPKILLDEATSAIDTSEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGF	603
DB	570	ARALIRNPKILLDMATSAIDNESEAMVQEVLSKIOGHGHTIISVAHRLSTVRAADTIIG	629
QY	604	DCGVIVYKGNHDELMKEGIFKLVMTQPRG-----NEIELENATGESKESDALEWSPKD	659
DB	630	EHGTAVERGTHEELLERKGVYFTLVTLQSQGNQALNEEDIKADDDMLARTFSRGVOD	689
QY	660	SGSLLIKRSTR---SIHAP--QGODRKLGTKED-----LNENVPVPSFWIRILKNS	707
DB	690	SLRASIRQRSKSQLSVLVHPEPLAVVDHKSSTYEEDRKDKDIPVQEEVAPVRRILKNS	749

Thu Nov 7 09:13:27 2002

```

270 ELKAYAGSVADEVISRTVAAGGKEKEYERKLNLFVFAQRWGIKRGIVMGFFTYM 329
305 FLIIYASYALAWYGFSLVL-SSEYITGOVLTVFFSVLIGAFSIGQASPSIFAFAFARGA 363
330 WCLIFFCYALAWYGSKLVEGEYSPGALVOIFLVIIGALNGLNAGSPCLFAFARGAA 389
364 AYEIKIIDKNPSIDSYSGSHKPNIKNLEFKNVHFSYPSKVEKVLKGLNKLKVGSGQ 423
390 ASSIFETIDRKPIIDCMSEGYKLERIKGEIEFHNTVPHYSPRPEVKILNLSWIKPGE 449
424 TVALVNSGCGKSTTVQLMORLYDPDGMVCDIGDIRINVRHLREITGVVSOBPVLFA 483
450 MTAIVGSPGAGKSTLQLHREFPTEGMVTVESHDIRSHIQWLKNQIGIYEQEPVLEF 509
484 TTAENIRYGRNVTWDETEKAVKEANAYDFIMKLPNKEDTTLVGERGAOLSGGQKORIAI 543
510 HTIAEKIRYGRDATMEDLLIOAKEANAYNFINDLPQFDTLVGEQGGQMSGGQKORVAI 569
544 ARALVRNPKILLDEATSDALDSEAVVQVALDKARKGRITIVIAHRLSTVRNADVIAGF 603
570 ARALIRNPKILLDMATSDALDSEAVVQVALDKARKGRITIVIAHRLSTVRNADVIAGF 629
604 DDGVIVKGNHDELMKEGIVFLVMTOTRGNELELE- ---NATGESKSESDALEMSPKD 659
630 EHGAAVERGTEELERKGVFALVLOSORNOGQDENEKDATEIDPEKTFSRGNYQD 689
660 SGSLIKRRS-----TRRSIHAPQGDOKLGYKEDNENVPVPSFWRIL 703
690 SLRASLRQSRQSLSYLAHEPPMAVEDHKSTHEEDRKDLPAQED---IEPASVRRIR 745
704 KLNTEWPYFVVGIFCALINGLOPAFSLIFSRLIGITFRDEDPETKRONNSMFSVLFLV 763
746 KLNAPWPMYDMLGSMGAAGVAPLFAFLFSQILGTFSL-PKKEORSGIINGICLLFVT 804
764 LGITSFTFFLQGTGFKAGBILTKRLRYMVFRLSMRQDYSWFDPPKNTTGALTTRLAND 823
805 LGCVSFTFFLQGTGFKAGBILTKRLRYMVFRLSMRQDYSWFDPPKNTTGALTTRLAND 864
824 AAQVKGATGSLAVITONIANLGTGIIISLYGWLILLALLAIVPIAIGVVMKMLSG 883
865 ASQVQAGTSGIOMMVNSFTNTVAMIAFLFSKLLGLIVCFPPFLLSALSGAJQTKMLTG 924
884 QALDKKELEGAGKATAEIENFTVSVLSLTREORFEYMYAQSLQVYRNSLRKAHIFGV 943
925 FASRDQALEKAGQITSSALSNIRTVAGIGKERFETFEAELEKPKYKMAIKKANVYGLC 984
944 FSITQAMFYSGAGFRGAYLVANEFMNFODVLLVFAIVFGAMVQGVSSFADPYAKA 1003
985 FGFSQITFIANSASYRGYGLISNEGLHFSYVFRVISAVALSATALGRASSYTPSYAKA 1044
1004 KYSAAHVIMIEKSPIDSYSPHGLKPNLTGNTVFNVEVFNYPTRPDIPVQLGSLSEVK 1063
1045 KISAARFFQLDRPPIINVYSAGEKWNFOGKIDFVDCKTTYFSRDPDIQVLNGLSVMS 1104
1064 KQOTLALVSGCGKSTVVQLLERYDPLAGSVLIDGKEIKHLNVQWLRAHLGIVSOBPI 1123
1105 PRQTLAFVSGSGKSTQLLERYDPDHGKVMIDGHSRKNVQFRLSNIGIVSOBPV 1164
1124 LFDCAENIAYGDNRSRVVSHIEIVQAAKEANIHFIEFLPEKYNTRYGDKGTOLSGGOK 1183
1165 LEACSIKDNKYGNDTQEIPEMRIIAAKAQVDFVMSLPEKYTEVNGSGQSLRGEK 1224
1184 ORTATARALVRPHILLDEATSDALDSEKVVQVALDKAREGRTCIIVIAHRLSTIONAD 1243
1225 ORTATARALVRPHILLDEATSDALDSEKVVQVALDKAREGRTCIIVIAHRLSTIONAD 1284
1244 LIYVFGQVKBEHGTQHQLAQKGYFMSV 1274
1285 IIAVMSQGVMEKGTHEELMVQKGYKLV 1315

```

RESULT 13

```

AB11_MOUSE
ID AB11_MOUSE STANDARD; PRT; 1321 AA.
AC Q9QY30: Q9QZE8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11)
DE (Sister of P-glycoprotein).
GN ABCB11 OR BSEP OR SPGP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RC MEDLINE=20076398; PubMed=10607905;
RX Green R.M., Hoda F., Ward K.L.;
RT "Molecular cloning and characterization of the murine bile salt export
RT pump.";
RL Gene 241:117-123(2000).
RN [2]
RP SEQUENCE OF 463-635 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RA Salkar R., Suchy F.J., Ananthanarayanan M.;
RT "Molecular cloning of mouse liver bile salt export pump (bsep).";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
CC INTO THE CANALICULUS OF HEPATOCYTES.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
CC MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
CC SITU.
CC -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF139303; AAF14372.1; -.
CC EMBL: AF186585; AAD56419.1; -.
CC MGI: MGI:1351619; Abcb11.
CC InterPro: IPR003593; AAA.
CC InterPro: IPR001140; ABC_transporter_tmem.
CC InterPro: IPR003439; ABC_transporter.
CC InterPro: IPR001687; ATP_GTP_A.
CC Pfam: PF00664; ABC_membrane; 2.
CC Pfam: PF00005; ABC_tran; 2.
CC SMART: SM00382; AAA; 2.
CC PROSITE: PS00211; ABC_TRANSPORTER; 1.
CC ATP-binding; Transmembrane; Transport.
KW DOMAIN 1 62 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 63 83 POTENTIAL.
FT DOMAIN 84 147 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 169 215 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 216 236 POTENTIAL.
FT TRANSMEM 237 240 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 262 319 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 320 340 POTENTIAL.
FT TRANSMEM 341 353 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 354 374 POTENTIAL.
FT TRANSMEM 375 755 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 756 776 POTENTIAL.

```

FT	DOMAIN	777	794	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	795	815	POTENTIAL.
FT	DOMAIN	816	869	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	870	890	POTENTIAL.
FT	TRANSMEM	891	911	POTENTIAL.
FT	DOMAIN	912	979	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	980	1000	POTENTIAL.
FT	DOMAIN	1001	1011	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1012	1032	POTENTIAL.
FT	DOMAIN	1033	1321	CYTOPLASMIC (POTENTIAL).
FT	NP_BIND	455	462	ATP (POTENTIAL).
FT	NP_BIND	1113	1120	ATP (POTENTIAL).
FT	CARBOHYD	109	109	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	122	122	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	481	481	L -> P (IN REF. 2).
FT	CONFLICT	633	633	T -> V (IN REF. 2).
FT	SEQUENCE	1321	AA; 146675 MW; 15B5EBF175D32967 CRC64;	
Query Match 49.7%; Score 3219.5; DB 1; Length 1321;				
Best Local Similarity 49.8%; Pred. No. 3.5e-176;				
Matches 644; Conservative 236; Mismatches 376; Indels 37; Gaps				
Qy	15	FWMGKSKKKEKE-KKPTVSTFAMFRYSNMLDRLYLVTGTMAAIIHGAALPLMLVFG	73	
Db	25	FHNDRKSLRQDKKGEARVGFELFRFSSSKDNWLMFPMGVCALLHGMAGPGLIIVFG	84	
Qy	74	NMTDSFANAGISR-----NKTFPVI---INESTNNTQHFINHLEEMTTYYAY	119	
Db	85	ILTDIFVEYDIERBELSPGKVCMMNTIYVINSFNQNNMTNGTSCGLVDINSEVIFSGI	144	
Qy	120	YSGIGAGVLVAAYIQVSWCLAAAGRLKLRKOFFHAIMROEIGWDFVDHVGELNRLTD	179	
Db	145	YAGVGVAVLLILGYFQIRLWVITGAROIRKMRKFYFRIMRMEIGWDFDCTSVGELNRSFD	204	
Qy	180	DVSKINEGIDGKMGFFQSIATFFGTVFTRGMKMLTLVLTAISPVGLSAAIWAKILS	239	
Db	205	DINKIDEATAQMALFQRLSTALSGLLGLGYRGMKMLTLVLTAISPVGLSAAIWAKILS	264	
Qy	240	SFTDKELLAYAKAGAEVLAARIVTIAFGGKKELEKRYNKNLEBEAGIGIKKAITANI	299	
Db	265	KFTELEKAYAKAGIADEVLSRITVAAGGENREVEREKNLMFAQRWGIWGMVNGF	324	
Qy	300	SIGAFLLIYAYALAFWYGTSLVL-SSEYITIGQVLTFFSVLIGAFSIGQASPSIEAFA	358	
Db	325	FTGYMMLIFFCYALAFWYGTSLVLDEGEYTPGLTIQIFLCVIAAMNIGNASSCLEIFS	384	
Qy	359	NARGAAAYETFKIIDNKPISDYSKSGHKPDNKGLEKFNHFSYPSRKEVKILGNLKL	418	
Db	385	TGCSAASSIFQIDROPVMDCMGSDGYKLDRIKGEIEFHNTFHYPSRPEVKILNLSMV	444	
Qy	419	VOSGOTVALVNGSGGKSTTVQLMORLYDPTDGMVCIDGODIRTNVRLHREITGVVSQE	478	
Db	445	IKPGETTAPVSGGAKSTALQIQRFYDPCGEGWTLDDHDIRSINRLWRDQIGIVEQE	504	
Qy	479	PVLFTATTENIRYGRENVTMDBIEKAYEANYDFIMKLPNKFTDLVGERGAQLSGGQK	538	
Db	505	PVLFTTATENIRLGREATMEDIVQAADANAYNFIMALPQDFTLVGEGGQMSGGQK	564	
Qy	539	QRIATARALVRNPKILLDEATSEAVVQVALDKARKGRTTIVIAHRLSTVRNAD	598	
Db	565	QRVAATARALIRPKILLDMATSLADNDESAKVQCALNKIQHRTIISVAHRLSTVRAD	624	
Qy	599	VIAGFDGVIYEKGNHDELMKEGIYFKLVMTQTRGNEIELENA-TGESKESDALEM--	655	
Db	625	VIIIGFEGHTAVERGTHTELLERKGYVFLVTLQSOEDNTHKETGIKGKDTTEGDTPERTF	684	
Qy	656	---SPKDSGSLIKRSTR-----SIHAPQGDQKRLGTGKEDLNENVPVSEF	700	
Db	685	SRGSYQDSLRSIRQSKSLSHLSHEPPLAIGDHKHSYEDRK--DNDVLVEEVEPAPVR	742	
Qy	701	RLKLNSTEWPFYVVGIFCAINGLQPAFSIIFSRILIGITFRDEDEPTRQNSMNFVSL	760	

DB	743	RILKYNISEWPIILVGCALCAINGAVTPIYSLFLSQILKTFSL-VDEQORSEIYSMCLF	801	
QY	761	FLVLGIISFIFFLQGFEGKAGELLTKRLRYMVFESMLRQDWFDDPKNTTGALTTL	820	
DB	802	FVILGCVSLFTQFLOGYNFAKSGELLTKRLKFGFKAMLRQDQIGWFDLKNPNGLTTL	861	
QY	821	ANDAAQVKAIGSLAVITQNIANLGTGIIISLYGQMLTLLALLAIPIIAIAGVVEKMK	880	
DB	862	ATDASOVQGGATGSOVGMVNSFTNIFVAVLIAFLFNKLSLVISVFPFELALSGAVQTKM	921	
QY	881	LSGQALDKKLEKAGAKIATEENERTVYVSLTREOKFEYMYAQSLQVYRNSLRKAHIF	940	
DB	922	LTGFSQDKEILEKAGQITNEALNIRTVAGIGVEGRFKAPEVELEKSYKTAIRKANVY	981	
QY	941	GVSESIQAMMYFSYAGCFRFGAYLVANEFMFQDVLVLSAIVFGAMAVGVSSFADPY	1000	
DB	982	GLCYAFSGIGISFLANSAAARYGGLIVYEDLNFYVFRVSSIAMSATAVGRTFTPSY	1041	
QY	1001	AKAKVSAHVIMIEKSPILDSYSPHGLKPNLTLEGWNTFVFNVPRPDIPVLOGLSL	1060	
DB	1042	AKAKISAAERFQLDRKPPIDVYSGAGEKWDNFQGRIDFIDCKFTYPSRDPDIQVNLGLSV	1101	
QY	1061	EVKKGOTLALVSGSGCKSTVOLLERFYDPLAGSVLIDGKEIKHLNQWLAHGLIYVSQ	1120	
DB	1102	SVDPGQTLAFVSGSGCKSTSIQLLERFYDPOGTVMIDGHSKKNVQVFLRSNIGIYVSQ	1161	
QY	1121	EPILFDCSIAENIAYGDSNRVVSHEEIVQAQAEANTHIEIETLPEKYNTRVGDGKTQSLG	1180	
DB	1162	EPVLFDCSIMDKYGDNTKEISVERAIAAQAOLHDFVMSLPEKYETNVGIGQSLSR	1221	
QY	1181	GOKORTAIRALVRPHILLDEATSEKVVQEAALDKARGRTCIVIAHRLSTIQ	1240	
DB	1222	GEKORTAIRAIRVDRPKILLDEATSEKVTSEKTVQALDKARGRTCIVIAHRLSTIQ	1281	
QY	1241	NADLVVVFQNGKVEGHTHOQLAOKGIYFSMW	1273	
DB	1282	NSDIIAMSGVGVIEGTHKKLMDQKGYKLV	1314	
RESULT 14				
AB11_RAT				
ID	AB11_RAT	STANDARD;	PRT;	1321 AA.
AC	070127;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Bile salt export pump (ATP-binding cassette, sub-family B, member 11)			
DE	(Sister of P-glycoprotein).			
GN	ABCB11 OR BSEP OR SPGP.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;			
RX	MEDLINE=98212048; PubMed=9543551;			
RA	Gerloff T., Stieger B., Hagenbuch B., Madon J., Landmann L., Roth J.,			
RA	Hofmann A.F., Meier P.J.;			
RT	"The sister of P-glycoprotein represents the canalicular bile salt			
RT	export pump of mammalian liver.";			
RL	J. Biol. Chem. 273:10046-10050(1998).			
CC	- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS			
CC	INFO THE CANALICULUS OF HEPATOCYTES.			
CC	- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN			
CC	THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR			
CC	MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN			
CC	SITU.			
CC	- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,			
CC	EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN			
CC	ATP BINDING CASSETTE (ABC) DOMAIN.			


```
MDRI_CAEL
ID MDRI_CAEL STANDARD; PRT; 1321 AA.
AC P34712;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein A).
GN PGP-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=93085750; PubMed=1360540;
RA Lincke C.R., The I., van Groenigen M., Borst P.;
RT "The P-glycoprotein gene family of Caenorhabditis elegans. Cloning
RT and characterization of genomic and complementary DNA sequences.";
RL J. Mol. Biol. 228:701-711(1992).
RN [2]
RP TISSUE SPECIFICITY.
RC STRAIN=BRISTOL N2;
RX MEDLINE=93223702; PubMed=8096815;
RA Lincke C.R., Broeks A., The I., Plasterk H.A., Borst P.;
RT "The expression of two P-glycoprotein (pgp) genes in transgenic
RT Caenorhabditis elegans is confined to intestinal cells.";
RL EMBO J. 12:1615-1620(1993).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: INTESTINAL CELLS.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X65054; CA446190.1; -
DR PIR; S27337; S27337.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 77 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 144 164 POTENTIAL.
FT TRANSMEM 213 233 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 321 341 POTENTIAL.
FT TRANSMEM 350 370 POTENTIAL.
FT DOMAIN 371 753 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 754 774 POTENTIAL.
FT TRANSMEM 798 818 POTENTIAL.
FT TRANSMEM 874 894 POTENTIAL.
FT TRANSMEM 895 915 POTENTIAL.
FT TRANSMEM 978 998 POTENTIAL.
FT TRANSMEM 1017 1037 POTENTIAL.
FT DOMAIN 1038 1321 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 451 458 ATP (POTENTIAL).
FT NP_BIND 1112 1119 ATP (POTENTIAL).
FT SEQUENCE 1321 AA; 145074 MW; 6130AAF3B31A8FA9 CRC64;
Query Match 42.3%; Score 2741; DB 1; Length 1321;
Best Local Similarity 45.1%; Pred. No. 7.6e-149;
Matches 585; Conservative 219; Mismatches 435; Indels 58; Gaps 12;
QY 21 KSKKKKKKKPTVSTFAMFRYSNNWLDRLMLVGTWMAAIIHGAALPLMLVFGNMTDSFA 80
DB 47 KTRDAKEEVNKKVSPQLRYITTTLEKLLDGLTGLTAVITAGLPLMSILOGKVSQAFI 106
QY 81 NAGISRNKTFPVIINESITNTQHEI-----NHLEEMTYAYIYSGIGAGVLVAA 132
DB 107 NE-----QIVIN-----NGSTFLPTGQNTKTDDEHDVNMVVSAAAMTVGMWAAG 154
QY 133 IQVSFWCLAAAGROILKIRKOFFHAIMRQBIGFWDHVDGELNTRTDDVSKINEGIGDKI 192
DB 155 IIVTCLYVAEQMNNLRREFVKSILRQESIFSWFDTHNSGTLATKLFDLRLVRKVGEGDKI 214
QY 193 GMFFQSIATFFFTGFIIVGTGRGNKLTILVLAISPVLGLSAAIWAKTLSSTFTKELLAYAK 252
DB 215 GMAFQVLSQFTGFIIVAFTHSQWQLTFLVLAIVPIQALCGFAIAKMSFTFAIRETLRYAK 274
QY 253 GAVAEVLAARTVIAFGGQKKELERYNNKLEAKGIGIKKAITANISIGAFTLLIYASY 312
DB 275 GKYVEETISSIRTVWSNLGRLELERYSTAVEAKKAGVLKGLFGISFGAMQASNFISF 334
QY 313 ALAFWYGTSLVLSSEVTIGQVLTFFSVLIGAFSGQASPIEAFANAGAAEYFIKIID 372
DB 335 ALAFYIGVGVHDGSLNFGDMLTTFSSVMGSMALGLAGPQLAVLGTAGQAASGIYEVLD 394
QY 373 NKPSIDYSKSGHKPDNIKGNLEFKNVHFSPSRKEILKGLNKVQSGQVVALVNSG 432
DB 395 RKPVIDSSSKAGRKDMKIKGDTIVENVHFTYSPRPDVPILGRMNLRYNAGQVVALVSGG 454
QY 433 CKKSTVQLMQLRYDPTGMCVIGDQDITRTINVRHLREITGVVSQEPVLPATTIAENRY 492
DB 455 CKKSTIISLLRYDVLKGTIDGVDRDINLEFLRNKNAVVSQEPALFNCTEENISL 514
QY 493 GRENTVMDIEKAVKEANAYDFIMKLPNKFTLVGERGAQLSGGQKORIAARALVNP 552
DB 515 GREGITREMWAAKMAAEFIKTLPGNYTLVGRGTQLSGGQKORIAARALVNP 574
QY 553 ILLDEATSALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVIAGDDGVIVEK 612
DB 575 ILLDEATSALDAESEGIQVQALDKAARKGRTTIVIAHRLSTIRNADLIISCKNGQVVEG 634
QY 613 NHDELMKEGIVFKLVMTWTRGNEI-----LEWATGESKESDALEMSFKDSSSL 664
DB 635 DHRALMAQGLYDILVTAQTFTDAVDSAAEGKFSRENSVARQTSHEGLSRQASEMDIM 694
QY 665 IK-RRSTRSI-HAPQ-----GQDRKGLTKEDLNENVPVSVFWRILKLNSTWPY-- 712
DB 695 NVRSTIGSITNGPVIDEKEERIGKDALSLKQELEEN-----NAQNTNFEILIYHA 747
QY 713 -----FVVGIFCALINGLQAPASIFSRIGIIFTRDEDPETKRONSNMFSVLVLGI 766
DB 748 RPHALSIFIGMSTATIGGIYPTYSVFTSPMNVFA--GNPADFLSQHFWMFLVLA 805
QY 767 ISFTIFFLQGTGFKAGELLTKRLRYMYERSMLRQDVSWEDDPKNTTGALTTRIANDAAQ 826
DB 806 AQGICSLMTFMGIASESLTRDNKLFNVLSSHQIHGFDFSDONASKGISTRLATDVPN 865
QY 837 VKGAIGSLAVITQNIANLGTGIIISLIYQWQLTLLLAIVPIIAIAGVVMKMLSGQAL 886
DB 866 LRTAIDFRFSIVITTLVSMVAGLGAFFYQWALLIALLPIVAGQVLRGRRTGKNV 925
QY 887 KDKKELEGAGKIATEATENFTVVSVLTREQFYMYAQSLQVYRNSLRKAHIFGVFSFI 946
DB 926 KSAEFADSGKIAIEATENRTVQALAREDTFYENCFEKLDIPKKAIEAFIQGLSYGC 985
QY 947 TQAMMYFSAGCRRFGAYLVANEFMFQD--VLLVPSAIVFGMAVGVSSAPDYAKAK 1004
DB 986 ASSVYLLNTCAIKWGLALIITPTMQPMRVLRVMYAITISTIGFATSFPEYAKAT 1045
QY 1005 VSAAHVIMIIIEKSPIDSYPHGLKPNLTLEGNTFNEVNFYTPRPDI PVQLGSLVYK 1064
```

Thu Nov 7 09:13:27 2002

Db 1046 FAGGIIFGMLRKISKIDSLSLAGEK-KKLYGKVIKKNVRFAYPERPEIEILKGLSFSVEP 1104
QY 1065 GOTLALVGSSGCGKSTVYQOLLERFYDPLAGSVLIDGKEIKHNLNQWMLRAHLGIYSQEPIL 1124
Db 1105 GOTLALVGPSCGCKSTVVALLEFYDPLGGEIFIDGSEIKTLNPEHTRSQIAIYSQEPTL 1164
QY 1125 FDCSIAENIAYGDNRSRVSHSEIYQAAKEANIHHFETLPEKYNTRVCGDKGTOLSGGQK 1184
Db 1165 FDCSIAENIYGLDPSSVTMAQVEEAARLANIHNFIAPLPEGFETRVGDRGTOLSGGQK 1224
QY 1185 RIAIARALVRQPHILLDLDEATSEKVVQVQALDKAREGRTCIVIAHRLSTIONADL 1244
Db 1225 RIAIARALVRNPKILLDLDEATSEKVVQVQALDKAREGRTCIVIAHRLNTVMNADC 1284
QY 1245 IVVFQNGKVKHGTHTQOLLAOKGIYFSMVSVQAGAKR 1281
Db 1285 IAVVSNGTIIIEKGTHTQOLMSEKGAAYKLTQKQWTEKK 1321

Search completed: November 6, 2002, 18:40:49
Job time : 17.4981 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:31:39 : Search time 29,3946 Seconds
(without alignments)
7539.009 Million cell updates/sec

Title: US-09-672-725C-27

Perfect score: 6473

Sequence: 1 MDPGGRKGSAGKFNWKGK.....LLAQKGIYFSWVQAGAKR 1281

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6425.5	99.3	1280	6	O46605
2	5682.5	87.8	1285	6	O02793
3	5641	87.1	1272	11	Q9JK64
4	5619	86.8	1272	11	Q924L6
5	5521.5	85.3	1163	6	Q9TSU2
6	5335	82.4	1169	11	O60502
7	4669.5	72.1	1288	13	Q93437
8	4444	68.7	1287	13	Q91586
9	3337	51.6	1348	13	Q90235
10	3221.5	49.8	1321	11	O88331
11	2866	44.3	1294	5	O45721
12	2832	43.8	1265	5	O01495
13	2825	43.6	851	13	Q9W693
14	2824.5	43.6	1275	5	O61301
15	2743	42.4	1321	5	Q21349
16	2695	41.6	1313	5	Q9V626

17	2648	40.9	1283	5	Q24393	Q24393 drosophila
18	2615.5	40.4	817	13	Q90WV6	Q90WV6 pseudopleur
19	2615.5	40.4	1285	10	O8LGX1	O8LGX1 oryza sativ
20	2612.5	40.4	1286	10	O80725	O80725 arabidopsis
21	2605.5	40.3	1292	10	Q9M109	Q9M109 arabidopsis
22	2601.5	40.2	1278	10	Q9FWX7	Q9FWX7 arabidopsis
23	2561	39.6	1279	5	Q9V616	Q9V616 drosophila
24	2533	39.1	1348	3	Q9Y8G1	Q9Y8G1 emericella
25	2532	39.1	1229	10	Q9SY12	Q9SY12 arabidopsis
26	2529	39.1	1229	10	O49749	O49749 arabidopsis
27	2527	39.0	1230	10	Q9SY13	Q9SY13 arabidopsis
28	2521	38.9	1289	10	Q94IH6	Q94IH6 coptis japo
29	2514	38.8	1349	3	O43121	O43121 aspergillus
30	2484	38.4	1229	10	Q9FWX8	Q9FWX8 arabidopsis
31	2466.5	38.1	1313	10	Q9ZRG2	Q9ZRG2 solanum tub
32	2466	38.1	1292	3	Q9C163	Q9C163 rhizomucor
33	2445	37.8	1408	3	O43140	O43140 cryptococcu
34	2437.5	37.7	1323	10	Q9MOM2	Q9MOM2 arabidopsis
35	2434.5	37.6	1248	10	Q9FHF1	Q9FHF1 arabidopsis
36	2426	37.5	1310	5	Q24852	Q24852 entamoeba h
37	2425	37.5	1286	10	Q9ZRW2	Q9ZRW2 arabidopsis
38	2417	37.3	1331	3	Q9HGT5	Q9HGT5 trichophyto
39	2416	37.3	1266	5	Q20331	Q20331 caenorhabdi
40	2403	37.1	1252	10	Q9LJX0	Q9LJX0 arabidopsis
41	2394	37.0	1320	5	Q9VRW2	Q9VRW2 drosophila
42	2391.5	36.9	1268	5	Q20335	Q20335 caenorhabdi
43	2386	36.9	1302	5	Q24851	Q24851 entamoeba h
44	2352	36.3	538	11	Q03982	Q03982 cricetus
45	2330.5	36.0	1222	10	Q9SVW6	Q9SVW6 arabidopsis

ALIGNMENTS

RESULT 1

O46605	PRELIMINARY;	PRT;	1280	AA.
ID	O46605			
AC	O46605;			
DT	01-JUN-1998 (TrEMBLrel. 06, Created)			
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	MULTIDRUG RESISTANCE P-GLYCOPROTEIN.			
GN	MDR1.			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
OX	NCBI_TaxID=9615;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLON;			
RA	Puel O., Lepage J.F., Alvinerie M., Galtier P., Pineau T.;			
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.			
CC	-!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.			
DR	EMBL; AF045016; AAC02113.1; -.			
DR	InterPro; IPR003593; AAA.			
DR	InterPro; IPR001140; ABC_transporter_tmem.			
DR	InterPro; IPR003439; ABC_transporter.			
DR	InterPro; IPR001687; ATP_GTP_A.			
DR	Pfam; PF00664; ABC_membrane; 2.			
DR	SMART; SM00382; AAA; 2.			
DR	PROSITE; PS00211; ABC_TRANSPORTER; 2.			
KW	ATP-binding; Transport.			
SQ	SEQUENCE 1280 AA; 141524 MW; 762DB5AFF4C73306 CRC64;			

Query Match 99.3%; Score 6425.5; DB 6; Length 1280;
Best Local Similarity 99.4%; Pred. No 0;
Matches 1273; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY	1	MDPEGGRKGSAGKFNWKGKSKKKKPTVSTFAMFRYSNWLDRLYMLVGTMAAII	60
Db	1	MDPEGGRKGSAGKFNWKGKSKKKKPTVSTFAMFRYSNWLDRLYMLVGTMAAII	59

QY	61	HGAALPLMLLVFGNMTDSFANAGISRNTFPVVI NESITNTQHFINHLEEMTYAYYY	120	Db	1140	VVSHEEIMQAKEANIHHFTETLPEKYNTRYGDKGQLSGGQRIARALVRQPHILL	1199
Db	60	HGAALPLMLLVFGNMTDSFANAGISRNTFPVVI NESITNTQHFINHLEEMTYAYYY	119	QY	1201	LDEATSDLTSESKVQVEALDKAREGTCIVIAHRLSTIQNADLI VVFQNGKVKHGTHQ	1260
QY	121	SGIGAGVLVAAYIQVSFWCLAAAGROILIRKQFFHAIMRQELGFWFDVHDVGELNTRLTDD	180	Db	1200	LDEATSDLTSESKVQVEALDKAREGTCIVIAHRLSTIQNADLI VVFQNGKVKHGTHQ	1259
Db	120	SGIGAGVLVAAYIQVSFWCLAAAGROILIRKQFFHAIMRQELGFWFDVHDVGELNTRLTDD	179	QY	1261	QLLAOKGIYFSMVSVOQAGAKR	1281
QY	181	VSKINEGIDKIMGFQSIATPFTGFI VGTFRGKMLTLVILAISVPLGSLAAIWAKILSS	240	Db	1260	QLLAOKGIYFSMVSVOQAGAKR	1280
Db	180	VSKINEGIDKIMGFQSIATPFTGFI VGTFRGKMLTLVILAISVPLGSLAAIWAKILSS	239	RESULT 2			
QY	241	FTDKELLAYAKAGAAVEVLAAIRVIAFGGOKKELERYNKNLEAKGIGIKKAITANIS	300	O02793		PRELIMINARY;	PRT; 1285 AA.
Db	240	FTDKELLAYAKAGAAVEVLAAIRVIAFGGOKKELERYNKNLEAKGIGIKKAITANIS	299	AC	O02793;		
QY	301	IGAFLIIYASALAFWYGTSLVLSSEYITIGQVLTVFFSVLIGAFSICQASPSIEAFANA	360	DT	01-JUL-1997 (Tremblrel. 04, Created)		
Db	300	IGAFLIIYASALAFWYGTSLVLSSEYITIGQVLTVFFSVLIGAFSICQASPSIEAFANA	359	DT	01-JUL-1997 (Tremblrel. 04, Last sequence update)		
QY	361	RGAAEIFIKIIDNKPSIDSYSGSHKPDNIKGNLEFKNVHFSYPSRKEVKILKGLNLKVQ	420	DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)		
Db	360	RGAAEIFIKIIDNKPSIDSYSGSHKPDNIKGNLEFKNVHFSYPSRKEVKILKGLNLKVQ	419	DE	MULTIDRUG RESISTANCE PROTEIN-1.		
QY	421	SGOTVALVNSGCGKSTTVOLMORLYDPTDGMVCIDGQDIRTINVRLHREITGVVSQBPV	480	GN	MDR1.		
Db	420	SGOTVALVNSGCGKSTTVOLMORLYDPTDGMVCIDGQDIRTINVRLHREITGVVSQBPV	479	OS	Ovis aries (Sheep).		
QY	481	LFATTIAENIRYGRENVMTDEIEKAVKEANAYDFIMKLPNKFTDLVGERGAOLSGGQKOR	540	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Db	480	LFATTIAENIRYGRENVMTDEIEKAVKEANAYDFIMKLPNKFTDLVGERGAOLSGGQKOR	539	OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;		
QY	541	IATARALVRNPKILLDDEATSDLTSEAVQVQALDKARKGTTTIVIAHRLSTVNRADVI	600	OC	Bovidae; Caprinae; Ovis.		
Db	540	IATARALVRNPKILLDDEATSDLTSEAVQVQALDKARKGTTTIVIAHRLSTVNRADVI	599	OX	NCBI_TaxID=9940;		
QY	601	AGDDGVIVEKGNHDELMKEGIFYFLVTMQTRNGEIELENATGESKSDALEMSPKDS	660	RN	SEQUENCE FROM N.A.		
Db	600	AGDDGVIVEKGNHDELMKEGIFYFLVTMQTRNGEIELENATGESKSDALEMSPKDS	659	RA	Longley M., Crawford A.M.;		
QY	661	GSSLIKRRTSRRIHAPOGQDRKLGKEDLNENVPVSWRILKLNSTWPFVVGIFCA	720	RT	"Ovine mdr1 gene."		
Db	660	GSSLIKRRTSRRIHAPOGQDRKLGKEDLNENVPVSWRILKLNSTWPFVVGIFCA	719	RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.		
QY	721	IINGGLQAPFISIRIIGITRDEDPETKRONSNFSLVLVLGIISITITFFLQGTFTG	780	CC	-1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.		
Db	720	IINGGLQAPFISIRIIGITRDEDPETKRONSNFSLVLVLGIISITITFFLQGTFTG	779	DR	EMBL; U78609; AAB58489.1; "		
QY	781	KAGEILTRLRYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSRLAVITQ	840	DR	InterPro; IPR003593; AAA.		
Db	780	KAGEILTRLRYMVFRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSRLAVITQ	839	DR	InterPro; IPR001140; ABC transporter_tmem.		
QY	841	NIANLGTGIIISLYGHQLTLLALLAIPITAIAGVEMKMLSGOALKDKKELEGACKIAT	900	DR	InterPro; IPR003439; ABC transporter.		
Db	840	NIANLGTGIIISLYGHQLTLLALLAIPITAIAGVEMKMLSGOALKDKKELEGACKIAT	899	DR	InterPro; IPR001687; ATP_GTP-A.		
QY	901	EAIENFTVSVLSTREQFEYMYAQSLQVYPYRNSLRKAHIFGVSFSTQAMMYFSYAGCFR	960	DR	Pfam; PF00664; ABC_membrane; 2.		
Db	900	EAIENFTVSVLSTREQFEYMYAQSLQVYPYRNSLRKAHIFGVSFSTQAMMYFSYAGCFR	959	DR	SMART; SM00382; AAA; 2.		
QY	961	FGAYLVANEFNFDVLLVSAIVFGAMAVQVSSFPADYAKAKVSAAHVIMIEKSPLI	1020	DR	PROSITE; PS00211; ABC_TRANSPORTER; 2.		
Db	960	FGAYLVANEFNFDVLLVSAIVFGAMAVQVSSFPADYAKAKVSAAHVIMIEKSPLI	1019	KW	ATP-binding; Transport.		
QY	1021	DSYSPHGLKPTNLGNVTENVFNPYTRPDIPVLOGLSLVKKGTALVLSGGCGKST	1080	SQ	SEQUENCE 1285 AA; 142020 MW; 90153B617C44856F CRC64;		
Db	1020	DSYSPHGLKPTNLGNVTENVFNPYTRPDIPVLOGLSLVKKGTALVLSGGCGKST	1079	Query Match	87.8%; Score 5682.5; DB 6; Length 1285;		
QY	1081	VVQLLERYDPLAGSVLIDGKEIKHLNVQWLRHLGTVSQEPILFDCSIAENIAYGDNRSR	1140	Best Local Similarity	87.3%; Pred. No. 2.8e-318;		
Db	1080	VVQLLERYDPLAGSVLIDGKEIKHLNVQWLRHLGTVSQEPILFDCSIAENIAYGDNRSR	1139	Matches 1122; Conservative	80; Mismatches 78; Indels 5; Gaps 3;		
QY	1141	VVSHEEIVQAANEANIHHFTETLPEKYNTRYGDKGTQLSGGQRIARALVRQPHILL	1200	QY	1	MDPEGRKG-SAEKNFWMGKK--SKKKEKKKPTVSTFAMFRYSNWLDRMLVGMTM	56
				Db	1	MDLEGRNGRAGGNFLKRDKKRPFSSKKDEKKRPTVSTFTMFRYSNWLDRMLVGMTL	60
				QY	57	AAIHGAALPLMLLVFGNMTDSFANAGISRNTFPVVI NESITNTQHFINHLEEMTY	116
				Db	61	AAIHGAAGLPLMLLVFGDMTDSFAGAGNLGNITLSNISNTSTIDRTE-YGKKLEKEMTY	119
				QY	117	AYYISGIGAGVLVAAYIQVSFWCLAAAGROILIRKQFFHAIMRQELGFWFDVHDVGELNTR	176
				Db	120	AYYISGIGAGVLVAAYIQVSFWCLAAAGROILIRKQFFHAIMRQELGFWFDVHDVGELNTR	179
				QY	177	LTDDYSKINEGIDKIGMFQSIATPFTGFI VGTFRGKMLTLVILAISVPLGSLAAIWAK	236
				Db	180	LTNDYSKINEGIDKIGMFQSIATPFTGFI VGTFRGKMLTLVILAISVPLGSLAAIWAK	239
				QY	237	ILSSFTDKELLAYAKAGAAVEVLAAIRVIAFGGOKKELERYNKNLEAKGIGIKKAIT	296
				Db	240	ILSSFTDKELLAYAKAGAAVEVLAAIRVIAFGGOKKELERYNKNLEAKGIGIKKAIT	299
				QY	297	ANISGAAELLIIYASALAFWYGTSLVLSSEYITIGQVLTVFFSVLIGAFSICQASPSIEA	356
				Db	300	ANISGAAELLIIYASALAFWYGTSLVLSSEYITIGQVLTVFFSVLIGAFSICQASPSIEA	359
				QY	357	FANARGAAYEIFKIIDNKPSIDSYSGSHKPDNIKGNLEFKNVHFSYPSRKEVKILKGLN	416


```
Db 360 FANARGAAEYFKIIDNKPISIDSYNTGHPKDNKGNLERNVHFHPSRNEVKILKGLN 419
Qy 417 LKVGSGOTVALVGNCGGKSTTQOLMORLYDPTDGMVCDIGQDIRINVRHRLREITGVVS 476
Db 420 LKVGSGOTVALVGNCGGKSTTQOLMORLYDPTDGMVSDIGQDIRINVRHRLREITGVVS 479
Qy 477 QEPVLFAATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGG 536
Db 480 QEPVLFAATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGG 539
Qy 537 OKORIATARALVRNPKILLDEATSDALTESEAVVOALDKARKGRTTIVIAHRLSTVRN 596
Db 540 OKORIATARALVRNPKILLDEATSDALTESEAVVOALDKARKGRTTIVIAHRLSTVRN 599
Qy 597 ADVIAGFDGCVIVYKGNHDELMKEGIYFKLVMTQTRGNEIELENAATESKESDALEMS 656
Db 600 ADIAGLDGCVIVYKGNHDELMKEGIYFKLVMTQTRGNEIELENAATESKESDALEMS 659
Qy 657 PKDGSGLIKRRSTRIRSIHAPOQDRKLGTKEDLNVPPVFWRLKLNSTWEPYFVVG 716
Db 660 SODSRSLIRKSTRIRSIHAPOQDRKLGTKEDLNVPPVFWRLKLNSTWEPYFVVG 719
Qy 717 IFCAIINGLOPAPSIIFSRIGITFRDDEPTKQNSMPSVLFVLGLIISFIFFLQG 776
Db 720 VFCAIINGLOPAPSIIFSRIGITFRDDEPTKQNSMPSVLFVLGLIISFIFFLQG 779
Qy 777 FTGKAGEILTRKLRVWFRSMRDVDFDPPKNTGALTRLANDAAQVKGAGSLA 836
Db 780 FTGKAGEILTRKLRVWFRSMRDVDFDPPKNTGALTRLANDAAQVKGAGSLA 839
Qy 837 VITQNTANIIGTIIISLYGWQLTLLLAIVPIIAAGVWEMKMLSGQALDKKKELEG 896
Db 840 VITQNTANIIGTIIISLYGWQLTLLLAIVPIIAAGVWEMKMLSGQALDKKKELEG 899
Qy 897 KATEAIENFRVWSLITROKFEYMYAQSLOVPRNSLRKAIHFGVSPSITQAMMYSYA 956
Db 900 KATEAIENFRVWSLITROKFEYMYAQSLOVPRNSLRKAIHFGVSPSITQAMMYSYA 959
Qy 957 GCFREGAYLVANEFNFDVLVFAIVFGAMVGOVSEAPDYAKAKVSAHVIMLIEK 1016
Db 960 GCFREGAYLVANEFNFDVLVFAIVFGAMVGOVSEAPDYAKAKVSAHVIMLIEK 1019
Qy 1017 SPLIDSYHGLKPNLTGNTVFNENPNYPRDIPVLOGLSLVKKGOTLALVSGSG 1076
Db 1020 IPLIDSYHGLKPNLTGNTVFNENPNYPRDIPVLOGLSLVKKGOTLALVSGSG 1079
Qy 1077 GKSTVVQLLERYDPLAGSLVDGKEIKHLNVOMLRAHLGLVSOEPIILFDCSIAENTAYG 1136
Db 1080 GKSTVVQLLERYDPLAGSLVDGKEIKHLNVOMLRAHLGLVSOEPIILFDCSIAENTAYG 1139
Qy 1137 DNSRVVSHBEIVQAAKEANIHFHTLPEKYNTRVGDKGTQSGQKORIAIARALVRP 1196
Db 1140 DNSRVVSHBEIVQAAKEANIHFHTLPEKYNTRVGDKGTQSGQKORIAIARALVRP 1199
Qy 1197 HILLDEATSDALTESEKVVQALDKAREGRTTIVIAHRLSTIONADLIIVFQNGVKEH 1256
Db 1200 HILLDEATSDALTESEKVVQALDKAREGRTTIVIAHRLSTIONADLIIVFQNGVKEH 1259
Qy 1257 GTHQOLLAQKGIYFSWVSQAGAKR 1281
Db 1260 GTHQOLLAQKGIYFSWVSQAGAKR 1284

RESULT 3
Q9JK64 ID Q9JK64 PRELIMINARY; PRT; 1272 AA.
AC Q9JK64;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DI 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DR MULTIDRUG RESISTANCE PROTEIN 1A.
GN PGY1.
```

```
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=LIVER;
RA Hooiveld G.J.E.J., Wilms J.W.J., Hagenbuch B., Jansen P.L.M.,
RA Meijer D.K.F., Muller M.;
RT "Cloning and functional characterization of the rat multidrug
RL resistance protein Mrda.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AF257746; AAF69007.1; -.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC_transporter_tm.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane; 2.
DR Pfam: PF00005; ABC_tran; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1272 AA; 140326 MW; B5C1DF89DFC986C5 CRC64;
```

```
Query Match 87.1%; Score 5641; DB 11; Length 1272;
Best Local Similarity 86.7%; Pred. No. 5.2e-316;
Matches 1110; Conservative 78; Mismatches 83; Indels 10; Gaps 2;
```

```
Qy 1 MDPEGGKSAEKNFWMGKSKKKEKKKPTVSTFAFRYSNWLDRLYMLVGTMAII 60
Db 1 MELEDNLGRADKNRSMGKKS-KKEKKKPAVSVLTFMRYAGWLDLRFYMLLGLTAAII 59
Qy 61 HGAALPLMLLVFGNMTDSFANAGISRNKTFPVIIINESITNTQFHINLEEMTYAII 120
Db 60 HGIALPLMLLVFGNMTDSFANVGNRSMF-----YNATDIYAKLEDEMTYAI 110
Qy 121 SGIGAGVLVAAYIQSVFCLAGROILKIRKQFFHAIMRQIGWDFVHDVGEINLRTD 180
Db 111 TGIGAGVLIVAYIQSVLCLAGROILKIRKQFFHAIMRQIGWDFVHDVGEINLRTD 170
Qy 181 VSKINEGIDKIGMFFQSIATFFFTGIVGTRGWKLTIVLAIISPVLGLSAIWAII 240
Db 171 VSKINEGIDKIGMFFQSIATFFFTGIVGTRGWKLTIVLAIISPVLGLSAIWAII 230
Qy 241 FTDKELLAYAKAGAAEVLAIRTVIAFGQKKELERYNNLEAKRLGKKAITANIS 300
Db 231 FTDKELQAYAKAGAAEVLAIRTVIAFGQKKELERYNNLEAKRLGKKAITANIS 290
Qy 301 IGAAFLIYASVALAFWGTSLVLSSEVTIGOVLTVFVSLIGAFSIGQASPIEAFANA 360
Db 291 MGAALFLLIYASVALAFWGTSLVLSSEVTIGOVLTVFVSLIGAFSIGQASPIEAFANA 350
Qy 361 RGAAYEIFKIIDNKPISIDSYNTGHPKDNKGNLERNVHFHPSRNEVKILKGLN 420
Db 351 RGAAYEIFKIIDNKPISIDSYNTGHPKDNKGNLERNVHFHPSRNEVKILKGLN 410
Qy 421 SGOTVALVGNCGGKSTTQOLMORLYDPTDGMVSDIGQDIRINVRHRLREITGV 480
Db 411 SGOTVALVGNCGGKSTTQOLMORLYDPTDGMVSDIGQDIRINVRHRLREITGV 470
Qy 481 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSG 540
Db 471 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSG 530
Qy 541 IAIARALVRNPKILLDEATSDALTESEAVVOALDKARKGRTTIVIAHRLSTVRN 600
Db 531 IAIARALVRNPKILLDEATSDALTESEAVVOALDKARKGRTTIVIAHRLSTVRN 590
Qy 601 AGFDGGVIVEQGNHDELMREKGIYFKLVMTQTRGNEIELENAATESKESDALE 660
Db 591 AGFDGGVIVEQGNHDELMREKGIYFKLVMTQTRGNEIELENAATESKESDALE 650
```



```

OX NCBI_TaxID=10031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=91154265; PubMed=1671863;
RA Devine S.E., Hussain A., Davide J.P., Melera P.W.;
RT "Full length and alternatively spliced pgp-1 transcripts in multidrug
RT resistant Chinese hamster lung cells.";
RL J. Biol. Chem. 266:4545-4555(1991).
CC 1- SIMILARITY, BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: M59234; AAA37005.1; -.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC_transporter_tmern.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane; 2.
DR Pfam: PF00005; ABC_tran; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR ATP-binding; Transport.
KW ABC_TRANSPORTER.
SQ SEQUENCE 1169 AA; 128939 MW; 72E25B7CE29DC185 CRC64;

Query Match 82.4%; Score 5335; DB 11; Length 1169;
Best Local Similarity 89.2%; Pred. No. 1.9e-298;
Matches 1043; Conservative 70; Mismatches 56; Indels 0; Gaps 0;

QY 113 MTTVAYTYSIGAGLVLAAYIQVSWCLAAAGROILKIRKOFFHAIMRQEIWDFVHDVGE 172
DB 1 MTTVAYTYSIGAGLVLAAYIQVSWCLAAAGROILKIRKOFFHAIMRQEIWDFVHDVGE 60
QY 173 LNTRLTDDVSKINIGIGDKTGMFQSTATFTFTGFTGKWLTLVLAISVPLGLSAA 232
DB 1 LNTRLTDDVSKINIGIGDKTGMFQSTATFTFTGFTGKWLTLVLAISVPLGLSAG 120
QY 233 IWAKILSFDTKELAYAKAGAAVEVLAARITVAFGGQKKELEKYNKLEAKGIGK 292
DB 1 IWAKILSFDTKELAYAKAGAAVEVLAARITVAFGGQKKELEKYNKLEAKRGLG 180
QY 293 KAITANTISGAFLLLIVASALAFWYGTSLVSEYITIGQVLVFFSVLIGAFSIGOASP 352
DB 1 KAITANTISGAFLLLIVASALAFWYGTSLVSEYITIGQVLVFFSVLIGAFSIGOASP 240
QY 353 SIEAFANAGAAEYEFKIDNKPIDSYSGSKHPDNKGNLEPKNVHFSYPSRKEYKIL 412
DB 1 NIEAFANAGAAEYEFKIDNKPIDSYSGSKHPDNKGNLEPKNVHFSYPSRKEYKIL 300
QY 413 KGLNLKVGSGTVALVNGSGCGKSTTVQLMORLYDPTDGVAVCIDQDIRTNVRLREIT 472
DB 1 KGLNLKVGSGTVALVNGSGCGKSTTVQLMORLYDPTDGVAVCIDQDIRTNVRLREI 360
QY 473 GVVSGEPVLFTTAENTRYRGENTWMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQ 532
DB 1 GVVSGEPVLFTTAENTRYRGENTWMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQ 420
QY 533 LSGGQKQRIATARALVRNPKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLS 592
DB 1 LSGGQKQRIATARALVRNPKILLDEATSAIDTSEAVVQVALDKARKGRTTIVIAHRLS 480
QY 593 TVRNADVTAGDDGVIVKGNHDELMKFKYFLVMTQTRGNELEENATGESKSESDA 652
DB 1 TVRNADVTAGDDGVIVKGNHDELMKFKYFLVMTQTRGNELEENATGESKSESDA 540
QY 653 LEMSPKDSGLIKRRSTRRSTHAPOGQDRKLGTKEDLNENVPVSWFRILKLNSTEWPY 712
DB 1 LEMSPKDSGLIKRRSTRRSTHAPOGQDRKLGTKEDLNENVPVSWFRILKLNSTEWPY 600
QY 713 FVVGIFCAINGLOPAPSIIFSRIGITFTRDEDPETKRONSWFSLVFLVIGISFTTF 772
DB 1 FVVGIFCAINGLOPAPSIIFSRIGITFTRDEDPETKRONSWFSLVFLVIGISFTTF 660
QY 773 FLQGFTEGKAGELLTKRLRYMVFERSMLRQDVSWFDDPKNTTGALTTRLANDAAQVGAIG 832
DB 1 FLQGFTEGKAGELLTKRLRYMVFERSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGATG 720
```

```

QY 833 SRLAVITONTANLGTGIIISLIYGMWLTLLLLLAIVPIIAIAGVVMKMLSGQALKDKKEL 892
DB 721 ARLAIVITONTANLGTGIIISLIYGMWLTLLLLLAIVPIIAIAGVVMKMLSGQALKDKKEL 780
QY 893 EGAGKIATEALENFRVTVSLTREOKFEYMAQSLQVQPYRNSLRKAHIFGVFSFTQAMMY 952
DB 781 EGSGKIATEALENFRVTVSLTREOKFENMAQSLQIPYRNALKAHVFGTFTSFTQAMMY 840
QY 953 FSYAGCFRFGAYLVANEFMFQDVLVLSAIVFCAMAGQVSSFPADYAKAKVSAAHVIM 1012
DB 841 FSYAACFRFGAYLVARELMTFENVLLVLSAIVFCAMAGQVSSFPADYAKAKVSAASHIM 900
QY 1013 IIEKSPILDSYSPHGLKPNTELEGNVTNEVFNFTPRDIPVLOGLSLEVKKGOTLALVG 1072
DB 901 IIEKVPISDSTGGLKPNTELEGNVKNEVFNFTPRDIPVLOGLSLEVKKGOTLALVG 960
QY 1073 SSGCGKSTVQVLLERFYDPLAGSVLDGKEIKHLNVQWLAHLGIVSQEPLTDCSTAEIN 1132
DB 961 SSGCGKSTVQVLLERFYDPMAGTVFLDGKEVNLQVWLAHLGIVSQEPLTDCSTAEIN 1020
QY 1133 TAYGNSRVVSHETVQAAKEANTHHTETLPEKYNTRVGDGTQLSGGQKQRIATARAL 1192
DB 1021 TAYGNSRVVSHETVQAAKEANTHHTETLPEKYNTRVGDGTQLSGGQKQRIATARAL 1080
QY 1193 VRQPHILLDEATSAIDTSEKVVQVQALDKAREGRTCTIVIAHRLSTONADLIYVFQNGK 1252
DB 1081 VRQPHILLDEATSAIDTSEKVVQVQALDKAREGRTCTIVIAHRLSTONADLIYVFQNGK 1140
QY 1253 VKEHGHQOQLLAQKGIYFSWVSQVQAKR 1281
DB 1141 VKEHGHQOQLLAQKGIYFSWVSQVQAKR 1169
RESULT 7
O93437 PRELIMINARY; PRT; 1288 AA.
AC O93437;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ABC TRANSPORTER PROTEIN.
GN CMDrl.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RX MEDLINE=99209805; PubMed=10195430;
RA Edelman H.M.L., Duchek P., Rosenthal F.E., Foeger N., Glackin C.,
RA Kane S.E., Kuchler K.;
RT "Cmdrl, a chicken P-glycoprotein, confers multidrug resistance and
RL interacts with Estradiol.";
RL Biol. Chem. 380:231-241(1999).
DR EMBL: AJ009799; CAA08835.1; -.
DR HSSP: P13569; INDB.
FT CHAIN 2 1288 ABC_TRANSPORTER_PROTEIN.
SQ SEQUENCE 1288 AA; 141917 MW; CB258A5F2826DB6C CRC64;

Query Match 72.1%; Score 4669.5; DB 13; Length 1288;
Best Local Similarity 71.3%; Pred. No. 4.2e-260;
Matches 920; Conservative 161; Mismatches 165; Indels 45; Gaps 12;

QY 2 DPEGGRKSAEKNFWMKMKKKKKKKKKKP-TVSTFAMFRYSNWLDRMLVGTMAAII 60
DB 27 DPEDEKKG-----KKKGGKPKQMVSPALAFRYSCTDKLLMIFGSLAIA 71
QY 61 HGAALPLMLVFGNMTDSFANAGISRNKTFPVIINNESITNNQFHINLHEEETTYAYY 120
DB 72 HTSLPIAMTIFGDMTDSFTVSGMTN-----ITGNSSGLNASSADVFNKLEEMTRYAYY 126
```

QY 121 SGIGAGVLAAAYIOVSFWCLAAAGROILKIRKQFFHAIMRQEIFHDFVHDVGEINLRLTDD 180
DB 127 SAIAAANVLVAAYIOTSFTWLAAGQVKKIRKREFFHAIMRQEIFHDFVHDVGEINLRLTDD 186
QY 181 VSKINEGIDKGIFGFIOSIAFTFTGFTVGTGRWKLTVLILASPVLGLSAAIAWAKILSS 240
DB 187 VSKINEGIDKGIFGLIQISFTFTLGTGFIWGKLTVLILASPVLGLSAAIAWAKILTA 246
QY 241 FTDKELAYAKAGAAVEEVLAARTVIAFGGQKKELERYNKNLEEAAGIGIKKAITANTS 300
DB 247 FTDKEQAYAKAGAAVEEVSARVTVIAFGGQKEIKRYHKNLEDAKIRKAITSNIS 306
QY 301 IGAAFLLIYASALAFWYGTSLVLSSEYTTIGQVLTFFSVLIGAFSIGOASPSIAFANA 360
DB 307 MGAFLLIYASALAFWYGTTLILANEYSIGNVLTFFSVLIGAFSIGOASPSIAFANA 366
QY 361 RGAAYEFKIIDNKPISDSYSGSKHPDNITKGNLEPKNVHFSYPSRKEVKILKGLNLYQ 420
DB 367 RGAAYAIENIDNEPEIDSYSDAGHKPDHIGKNEFQNVFFNPSRDPVEILKGLNLYN 426
QY 421 SGOATVALVNGSGGKSTTVMOLRYDPTDGMVICIDGDIRINVRHLREITGVVQBPV 480
DB 427 CGQTVALVNGSGGKSTTVOLIQRFDYDPKEGTITIDGDLKSLNRYLREITGVVQBPV 486
QY 481 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFTLTVGERGAOLSGGQKOR 540
DB 487 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFTVVGGERGAOLSGGQKOR 546
QY 541 IATARALVRNPKILLDEATSAIDTSEAVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
DB 547 IATARALVRNPKILLDEATSAIDTSEAVQVALDKARKGRTTIVIAHRLSTVRNADLI 606
QY 601 AGFDGIVGVEKGNHDELMKEKGYKFLVTMOTRGNIELEENATGESKESDALEMSPKDS 660
DB 607 AVFESGVITEQGNHLSOLIEBKGYIKLVNMQT-----IETEDPSS-KSEN---AVSVKRS 658
QY 661 GS-----SL-----IKRRSTRRIH---APQGDRLKGTKEDENNVPPVSEWRILKLNST 708
DB 659 GSQSNLDESILKELRGSTRSRMKPGEPNDTDER-GSSPD--EELPPVSLKMLKLNKN 715
QY 709 EWYFVVGIFCALTNGLOPAISIFSRILGIFTRDEDEPKRQNSMFSVLFLVLGIIS 768
DB 716 EWYFVAGTFCALVNGALQAPAFSVIFSEILGIFS-ETDQKVLREKSNLYSLFLALGIIS 774
QY 769 FITFFLQGTFFGAGELTKRLRYVMVFRSMLRQDYSWFDPKNTTGALTTRLANDAQVK 828
DB 775 FTFFVQGFAGFAGKAGEILTMKLRFAKMLEQDMANWFDPKNSTGALTTRLANDASQVK 834
QY 829 GAIGSLAVITONIANLGTGIIISLIYQWLTLLALLAIVPIIAIAGVYEMKMLSGQALKD 888
DB 835 GATGVRLALIAQNIANLGTGIIISLIYQWLTLLALLAIVPIIAIAGVYEMKMLAGHAKD 894
QY 889 KKELEGAKIATEATENRTVSTLTREOKFEYVYAQSLOVYRNSLRKAHFGVFSFISQ 948
DB 895 KLEEAAGKIAATEATENRTVASLTRERFELMGEHLVLYRNSVKKAHFGFCFSLQ 954
QY 949 AMMYTSAGCFRFGAYLVANEFMNPQDVLVFSALVFGAMAVGVSSFADYAKAKYSAA 1008
DB 955 AMMYTSAGCFRFGAYLVVNGHIEKTVLVSFVAVFGAMALGQTSFADYAKAKISAA 1014
QY 1009 HWIMIEKSPIDISVPHGLKNPTLEGNVTFNEVVFYTRPDIPVLQGLSLEYKKGQTL 1068
DB 1015 HLFLVFNRPVPTDSYREDGEPKPEFGGNTRIKOVKNYPNRPEVKILQGLNLAVERGETL 1074
QY 1069 ALWSSGGCKSTVOLLERFVDPLAGSVLIDGKEIKHLNVQWLRAHGLGVISQBPILFDCS 1128
DB 1075 ALWSSGGCKSTVOLLERFVDPLUSGEIVFDIDAKTLNTQWLSHIGVYSQBPILFDET 1134
QY 1129 TAENIAYGDNRSVSHIEIVAAKEANIHHPETLPEKYNTRVGDGKTQLSGGQKORIAI 1188
DB 1135 TAENIAYGDNRSVSHIEIIISAAKASHFSDLSLPEKYNTRVGDGKTQLSGGQKORIAI 1194

QY 1189 ARALVROPHILLDEATSAIDTSEKVVQBALDKAREGRTCTIVIAHRLSTIQNADLIVVF 1248
DB 1195 ARALIRKPOILLDEATSAIDTSEKIVQBALDKAREGRTCTIVIAHRLSTIQNADKIAVI 1254
QY 1249 ONGKVKEGHGTHOQLLAQKGIYFYSVSVQACA 1279
DB 1255 ONGKVIQOGTHOQLLAEKGFYISLVNVQSGS 1285
RESULT 8
Q91586 PRELIMINARY; PRT; 1287 AA.
AC Q91586
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN.
GN XEMDR.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95322451; PubMed=7599185;
RA Castillo G., Shen H.J., Horwitz S.B.;
RT "A homologue of the mammalian multidrug resistance gene (mdr) is
RL functionally expressed in the intestine of Xenopus laevis";
RL Blochim. Biophys. Acta 1262:113-123(1995).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: U17608; AAA75000.1; -.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC transporter_tmem.
DR InterPro: IPR003439; ABC transporter.
DR InterPro: IPR001687; ATP_Gmp_A.
DR Pfam: PF00664; ABC membrane; 2.
DR Pfam: PF00005; ABC_tran; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1287 AA; 141505 MW; 06E95073C5771415 CRC64;
Query Match 68.7%; Score 4444; DB 13; Length 1287;
Best Local Similarity 67.9%; Pred. No. 4e-247;
Matches 874; Conservative 171; Mismatches 209; Indels 34; Gaps 11;
QY 2 DPEGGRKSAEKNFWMGKSKKKEKPKTSTFAMFRYSNMLDRLMLVGTMAAIIH 61
DB 21 DPNSNSK--EKKGFF--SKFKKKKEKTEKPKVGVFTMFYSSSTDKMLLFGFIASLAH 76
QY 62 GAALPMLLVFGNMTDSPANAG--ISRNKTPFVINESITNTTOHFINHLEEMTTAYY 119
DB 77 GAALPMLLVFGEMTDSFNVNQVDTGNFTWESMINAS-----RELQGMVTVAYY 127
QY 120 YSGICAGVLAAAYIOVSFWCLAAAGROILKIRKQFFHAIMRQEIFHDFVHDVGEINLRLTDD 179
DB 128 YSGLGFGVWMLCAIYQISFWTUSAGHQIKKIRSNFFHVAFLRQEIFGWFINDAGELNRLTDD 187
QY 180 DYSKINEGIDKGIFGFIOSIAFTFTGFTVGTGRWKLTVLILASPVLGLSAAIAWAKILSS 239
DB 188 DYSKINEGIDKGIALMLLQSLTTLVYTGIFIGKWKLTWVGALSPINGLSAAIAWAKVLS 247
QY 240 SFTDKELAYAKAGAAVEEVLAARTVIAFGGQKKELERYNKNLEEAAGIGIKKAITANI 299
DB 248 APTNKLKAYAKAGAAVEEVLSSIRTVFAFGQNKKEIHRYEKKNLEDAKIRKAITANV 307
QY 300 STGAFLLIYASALAFWYGTSLVLSSEYTTIGQVLTFFSVLIGAFSIGOASPSIAFANA 359
DB 308 SIGFAFLMIYAAVSLAFWYGTTLIDGGYITIGSVLTFFFAVITGAFVAGTSNIEAFAN 367
QY 360 ARGAAEYEFKIIDNKPISDSYSGSKHPDNITKGNLEPKNVHFSYPSRKEVKILKGLNLYQ 419

Db	368	ARGAAYTIFNIIDNOPKIDSFKEGLKDPDKINGDIEFKNVIFTYPSRKDIQVLKGLNLNI	427
Qy	420	QSGOTVALVNGSGCKSTTVQLMORLYDPTDGMVCIQDQDINTINVRHLREITGVVSQEP	479
Db	428	PSGKTVALVSGSGCKSTTVQLIQFYDPEQGVITLDGQDIRSLNIRVLRITGVVSQEP	487
Qy	480	VLFATTIANIRYGRNVMTWDEIEKAKVEANAYDIMKLPKDFPLVGERGALSGGOKO	539
Db	488	ILFDTTIADNIRYGRNVDVTKIEERATKEANAYDIMKLPKLETLVGERGQLSGGOKO	547
Qy	540	RTAIARALVRNPKILLDDEATSEAVVOALDKARKGRTTIVIAHRLSTVRNADV	599
Db	548	RTAIARALVRNPKILLDDEATSEAVVOALDKARKGRTTIVIAHRLSTVRNADV	607
Qy	600	IAGFDGVIKVERGNHDELMKRGYFKLVMTOT---RGNIELENATGESKESDALEM	655
Db	608	IAGFDGVIKVERGNHDELMKRGYFKLVMTOT---RGNIELENATGESKESDALEM	660
Qy	656	SPKDSGSLIKRRSTRSTHA--POQODRKLGTKE-DLNNENPVVPSFWRILKLNSTWPY	712
Db	661	PVTHSNLVRKSSRNTKSKVPETEDKEVEDEEKKKEEGPPPVFFKVMKLNKPEWY	720
Qy	713	FVYGICFAINGLOPAFISIRIIGTFRDDEPETK-RQNSNMFSLVLVLGHISFIT	771
Db	721	FVVGVICAMINGATQAPAFIIFSRIGVFA--GPVSOQRSESSMYSLFLALGGVSFIT	777
Qy	772	FFLOGFTFGKAGBELTKRKYWFRSMLRODVSFDDPKNTTGALTURLANDAAQVKAI	831
Db	778	FFLOGFTFGKAGBELTKRKYWFRSMLRODVSFDDPKNTTGALTURLANDAAQVKAI	837
Qy	832	GSLAVITONIANLGTGIIISLYGQWLTLLLAIPITAIAGVVMKMLSGOALKDKKE	891
Db	838	GTRLLALQNVANLGTGIIISLYGQWLTLLLAIPITAIAGVVMKMLSGOALKDKKE	897
Qy	892	LEGAGKIATEAENFRVWSLTLREQFEFYWAQSLQVPYRNSLRKAHIFGVFSITQAMM	951
Db	898	LEKAGKISTDAVLNIRVSLTRERKPEAMYEKSLGEPYRNSIKKALHGLTYGLSQAH	957
Qy	952	YFSVAGCFR-FGAYLVANEFNPDVLLVFAIVFAGAMAGVQVSSPAPDYAKAKVSAAHV	1010
Db	958	VLCWCWFVSLGAYLVVEGLMKLDEFLVSSAIVLGNMAGLQTSFSPAPDYTKAMISAAHI	1017
Qy	1011	IMITLESPLIDSYPHGLKPNLTLEGNVTFNVEVFNYPTRPDIPVLQGLSLEVKKQTLAL	1070
Db	1018	FSLLERVPQIDSYSQDQEKPKNCNGVNVFKNVFNYPTRPDITVLOGLDISVRQGETLAL	1077
Qy	1071	VGSSGCKSTTVQLLRYDPLAGSLVDGKEIKHLNVQWLRHGLTVSOEPILEDCSIA	1130
Db	1078	VGSSGCKSTTVQLLRYDPLAGSLVDGKEIKHLNVQWLRHGLTVSOEPILEDCSIA	1137
Qy	1131	ENIAYGNSRVVSHIEIVQAAKEANIHHFTETLPEKYNTRVGRKQTLQSGQKQRIAIAR	1190
Db	1138	DNIAYGNNRKYVQEEIETAAKEANIHSFIESLTDKYNTRVGRKQTLQSGQKQRIAIAR	1197
Qy	1191	ALVRQPHILLDEATSEKVDQALDKAREGTCIVIAHRLSTVONADLVVQFN	1250
Db	1198	ALVRQPHILLDEATSEKVDQALDKAREGTCIVIAHRLSTVONADLVVQFN	1257
Qy	1251	GKVEHGHQOALLAQKGYFESVSVQAG 1278	
Db	1258	GKVEHGHQOALLAQKGYFESVSVQAG 1285	
RESULT 9			
Id	Q90235	PRELIMINARY; PRT; 1348 AA.	
AC	Q90235		
DT	01-DEC-2001	(Tremblrel. 19, Created)	
DT	01-DEC-2001	(Tremblrel. 19, Last sequence update)	
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)	
DE	BILE	SALT EXPORT PUMP.	
OS	Raja	erinaea (Little skate).	
OC	Eukaryota;	Metazoa;	Chordata;
		Craniata;	Vertebrata;
		Chondrichthyes;	

OC	Elasmobranchii; Squalae; Hypnosqualae; Pristiorajae; Batoidea;
OC	Rajiformes; Rajidae; Raja.
OX	NCBI_TaxID=7782;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=21340059; PubMed=11447010;
RA	Cai S.Y., Wang L., Ballatori N., Boyer J.L.;
RT	"Bile salt export pump is highly conserved during vertebrate evolution
RL	and its expression is inhibited by PFIC type II mutations.";
RRL	Am. J. Physiol. Gastrointest. Liver Physiol. 281:G316-G322(2001).
DR	EMBL; AF367243; AAK52958.1; -
SQ	SEQUENCE 1348 AA; 148648 MW; 81906F641654CFDD CRC64;
Query Match	51.6%; Score 3337; DB 13; Length 1348;
Best Local Similarity	50.3%; Pred. No. 2.2e-183;
Matches 659; Conservative 249; Mismatches 340; Indels 62; Gaps 9;	
Qy	22 SKKKE-----KKEKKPTVSTFAMFRYSNLDRLYMLVGTMAAIIHGAALPLMLVFGNM 75
Db	44 SKSKDDNENLSKKGRDKIGFFQLFERFASCYEVTLWFGVGCALLHGVQAPVALLVFGLL 103
Qy	76 TDSFANAGISRNKTFPVIINESITNTOHFIN-----HLEEMTTIYAY 118
Db	104 TDTFIAYDIELOELKDT--RKHCNNTIHWNGSEYLNQHNMTSCGILNIEKEMTFAY 161
Qy	119 YYSGITAGVLVAAYIQVSWFCLAAAGRLIKRKOFFHAIHQEIGWFDVHDVGEINLRLT 178
Db	162 YVYGICSVLLLVGYQICFWVTAARQTKIRKAYFRQIMRMEMGWDFCNSVGEINLRMS 221
Qy	179 DDVSKINEGIDKMGFPQSTATFTFTGTRGKLTIVLAIISPVLGISAATWAKIL 238
Db	222 DDINKINDAIADQVGIFQRFSTFVSGFLMGFVNGKLTIVIAVSPPLGLGAALMALSV 281
Qy	239 SSTDKELLAYAKAGAAVEVLAARTVIAFGQKKELERYKNLEAKGIGIKKAITAN 298
Db	282 ARLTGLGLKAYAKAGAAVEVLSIRTVAAFGSEKEDYDRDLVFAQRWIRKGMIMG 341
Qy	299 ISIGAFLIIVASALAFWYGTSLVL--SSEYTGIVLTIVFVSVLIGAFISGOASPSIRAF 357
Db	342 FFTGYVMIIFFCYALAFWYGTSLVIEQNEYPGGLQVFLVLAAMNLQASPCLEAF 401
Qy	358 ANARGAAVEIKIIDNPKSIDSYSKSHKPNKINLEKFNHVSYPKSRKEVILKGLNL 417
Db	402 ASGRGAARKIFETIDREPEIDCMSEGGYTLNKKVGDIEFHNVTFNPSDPVKTLDRLSM 461
Qy	418 KVQSGOTVALVNGSGCKSTTVOLMORLYDPTDGMVCIQDQDINTINVRHLREITGVWSQ 477
Db	462 VIRAGETTAFVGPSSGSKSTAVELIQRFYDPKQGMVITLDGHDHDIRSLNIQWLRSLIGIVEQ 521
Qy	478 EPVLFTATTAEINIRYGRNVMTWDEIEKAKVEANAYDFIMKLPKFDTLVGERCAQLSGGQ 537
Db	522 EPVLFTATTAEINIRYGRNVMTWDEIEKAKVEANAYDFIMDLPEKFTLVLVEGGGQMSGGO 581
Qy	538 KORIAIARALVRNPKILLDDEATSEAVVOALDKARKGRTTIVIAHRLSTVRNA 597
Db	582 KORIAIARALVRNPKILLDDEATSEAVVOALDKARKGRTTIVIAHRLSTVRNA 641
Qy	598 DVTAGDDGVYVEKGNHDELMKRGYFKLVMTQTRGNEIELENATGESKESDALEMS 657
Db	642 DVIIGFHBGAVRGRHAELEKRGIVYTLVLTQTQGEQALHEKA---RQVNGAIEDGA 697
Qy	658 KDSGSLIKRRSTRSTHAPOQODRK-----LGTK----- 687
Db	698 SEK-RQLIRGSSRASVSTRHORSRSQVSEVLSDLSPGDVAVAVRTPSISLGDEKDV 756
Qy	688 EDLNENPVVPSFWRILKLNSTWPEYVVGIFCAIINGLOPAFISIFSLIGIFTRDEDP 747
Db	757 EEEEEESTEPAPVSRILKYNVSEWYMLFGLSGLAAGVGNVPYIALLFSQILGTFSL-QNE 815
Qy	748 ETRQNSNMFSLVLVLGIIISFTTFFLQGTFTFKAGEILTKRLRYMVFRRSLRQDVSFWD 807
Db	816 EEKINQINAICLFFVVGVLVSFLTQFLQSYFFFAKSGELLTRRLKRLGFAQLRQEIWFDP 875

```
QY 808 DPKNTGALTTRLANDAAQVGAIGSRVAVITONIANLGTGIIISLIYGMQLTLLIAIV 867
Db 876 DRKNSPGTLTTRLATASQVQGTQIGMIVNSITNIGVSLIIAFYFSWKLIVILCFL 935
QY 868 PIIATAGVVENKMLSGQALKDKLEGAGKIATAEIENFRIVSVLREQFEYMAQSLQ 927
Db 936 PFLALTGALQARMUTGFANQDKEALEAAGQISSEALSNIRTIAGLAKEMFVOLFEAQL 995
QY 928 VYRNSLRKAHIFGVFSFIQAMMYFAGCFRFGAYLVANEFMNFQDVLVLSAIVFGA 987
Db 996 APYAAIAKANNVYGLCFGAQSVIFLANAASYREGGLVVAENLHRSIVFRVLSALVTS 1055
QY 998 MAVGOVSSFADYAKAVSAHVIMITEKSPIDISYSPHGLKPNTELEGNVTFNEVFNYP 1047
Db 1056 TALGRASSTYDPAKAKIAAAREFQLIDRAPKINIDSEGEKSNFRGDKFVECTFTY 1115
QY 1048 TRPDIPVLQGLSVKVGQTLALVGGSCGKSTVQVLLERFYDPLAGSVLIDGKEIKHLN 1107
Db 1116 TRPDFQVNLGSLVCMFGQILAFVGGSCGKSTVQVLLERFYDPLAGSVLIDGKEIKHLN 1175
QY 1108 VQWLRHAGLIVSQBPILFDCSIAENIAYGDSNRVWSHEETVQAAKENIHHFTETLPEKY 1167
Db 1176 TAFRLRIGIVSQBPVLFDSIEENIKYGDSNRQVSMDEVIAAKQAQLSHFVMSLPEKY 1235
QY 1168 NTRVGDKGTOLSGOKORIAIARALYRQPHILLDEATSAIDTESKVVQOALDKAREGR 1227
Db 1236 GTRVGAQGSQSLSGOKORIAIARVDPKILLDEATSAIDTESKIVQDALDEARKGR 1295
QY 1228 TCIVIAHRLSTIONADLIVVFQNGKVEKHTHOOILAAQKGYFSVMVSQA 1277
Db 1296 TCIVIAHRLSTIONSIIAIVMSRGIIEQSGHNLMAAKGAYIKLVTTGA 1345

RESULT 10
O88331 ID O88331 PRELIMINARY; PRT: 1321 AA.
AC O88331;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE P-GLYCOPROTEIN SISTER.
DE P-GLYCOPROTEIN SISTER.
GN SPGP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RC TISSUE=LIVER;
RA Childs S.J., Yeh R.L., Hui D., Ling V.;
RT "Taxol resistance mediated by the liver-specific sister gene of P-
RT glycoprotein.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF010597; AAC24753.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmam.
DR InterPro; IPR001687; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 1321 AA; 146286 MW; 27F67EC36608D0F CRC64;

Query Match 49.8%; Score 3221.5; DB 11; Length 1321;
Best Local Similarity 50.5%; Pred. No. 9.3e-177;
Matches 650; Conservative 233; Mismatches 369; Indels 35; Gaps 11;

QY 20 KSKKKKEK-KEKKPTVSTFAMFYSNNLDRLYLVLGVTWAAIIHGAALPLMMLVFGNMTDS 78
Db 30 KKSRLQDKMKEGDIRVGFELFRFSSSKDIWLMGMGVCALLHGMAGPGIILIFGINTDI 89
```

```
QY 79 FANAGISR-----NKTFFVINES-----ITNTOHFHNLHEEEMTYAYYSGI 123
Db 90 FIKYDIEROELEIFPKACVNNNTI~VINSNFQNMNTVCGGLVDISEMKEFSIYAGV 148
QY 124 GAGVLVAAYIOVSWPCLAAAGROILKIRKQPFHATMRQEIQWFDVHDVGTENLTLDVSK 183
Db 149 GMTVLLICYFOIRLWVITGARQIRMRKTIYFRIRMEIGWFDCTSVGELNSRFADIEK 208
QY 184 INEGIDKIGMFOSIATFTFTGIVGTGRKLTVLTAISPLVGLSAAIWAALISFTD 243
Db 209 INDIAIOLAHLFORMSTAMCGLLGLFYRGWKLTLVLAIVSLIGAAVIGLSIAKFT 268
QY 244 KELLAYAKAGAVAEVLAARTVIAFGQKKELEERNKNLEAKGIGIKKAITANISGA 303
Db 269 LELKAYAKAGSIADIEVLSIRTTAAAFGENKEVEREKNLVFAQRWIKGVMVFFTCY 328
QY 304 AFLIYASIALAFWYGTSLVL~SSEYTIQVLVFFSVLIGAFSIGOASPSIFAEANARG 362
Db 329 MWCLIFFCYALAFWYGTSLVDEEYTPGLVOIFLCVILAAAMNIGHASCSLEIFSTG 388
QY 363 AAYEIFIKIIDNKPSIDYSKSGHKPNIKGNLEKFNHFSYPSRKEVKILGNLKVQSG 422
Db 389 AATNIFQTIQRPVIDCMGDKYKLDRIKEIEFHNWTFHYPSRPDKILDNLNVKPG 448
QY 423 QTVALVNSCGGKSTTVQLMORLYDPTDGMVCIQDQDITINVRHLREITGVVSOEPLV 482
Db 449 ETALVSSGAGKSTALQILQRFYDPCGMVTLDDGHDIRSLNIRWLDRIGIVEQPLV 508
QY 483 ATTIAENIRYGRNVMTDETEKAVKEANAYDFIMKPNKFDTLVGERGAOLSGGOKORTA 542
Db 509 STTIAENIRFREDATMEDIVQAAKANAYNFMALPQOFTLVGEGGQSGGOKQVRA 568
QY 543 IARALVRNPKILLDEATSAIDTESAVVQVALDKARKGRITTVIAHRLSTVRNADVIAG 602
Db 569 IARALIRNPKILLDMATSAIDNESARVQEAALNKIQHGTITISVAHRLSTVRAADVI 628
QY 603 FDDGVIVEKGNHDELMEKGIYKLVMTQTRGNEIELENA~TGESKSSDALEM-----S 656
Db 629 FEHGVAVERTHEELLERKGVFMVLVLOSQGDNAHKETSIINGKDATEGGLTFRSGS 688
QY 657 PKDSGSLIKRRSTR~--SIHAP--QGDRKLGTKE~----LNENVPVPSFWRILKN 706
Db 689 YRDSLRAIRQSKSLSLLTHDPPLAVADHKSYSKSDKNDVIVEEVEPAPVRILKYN 748
QY 707 STEWPFVVGIFCAIINGLOPAFSLIIFSRITGIFTRDEDPETKQNSNMFVSLFLVGI 766
Db 749 IPEWHYILVGLSAAINGAVTPIYSLFSQLLGTSL~LDKEQQQRSETHSMCLFVLIG 807
QY 767 ISPTFFLGQFTGKAGEILTAKRLRVWVFRSMLRODVSWFDDPKNTTCALTTRLANDAA 826
Db 808 VSIPTFOLOGYFAKSGELLTRLRKFGKAMLGQDQIGWFDLDRNPGVLTTRLATDASQ 867
QY 827 VKGAGSRVAVITONIANLGTGIIISLIYGMQLTLLIAIVPIIAIAGVVMKMLSGOAL 886
Db 868 VOGATGSGVMVNSFTNIIAALLIAFFFSWKLSLIITIFFPFLALSGAVTKMLTGAS 927
QY 887 KDKKELEGAGKIATAEIENFRIVSVLREQFEYMAQSLQVYRNSLRKAHIFGVFSFI 946
Db 928 QDKQALEKAGQITSEALSNIRTVAGIVEGREFIKAFEVELOTSYKTAVRANIYGLGFAF 987
QY 947 TQAMMYFVAGCFRFGAYLVANEFMNFQDVLVLSAIVFGAMAVGOVSSFADYAKAVKS 1006
Db 988 SOGIAFLANSAAIRYGGYLLIAYEGLGFHSVFRVSVSVLSTAVGRFTSTPSTAKAKIS 1047
QY 1007 AAHVIMITEKSPIDISYSPHGLKPNTELEGNVTFNEVFNFTPRDIPVLQGLSVLKKG 1066
Db 1048 AARFQQLIDRKPPIINVYSEAGEKWDNFOGIDFIDCKFTYPSRPDIQVNLGSLVSNPG 1107
QY 1067 TLIALVSSGCGKSTVQVLLERFYDPLAGSVLIDGKEIKHLNVQWLRHAGLIVSQEPLD 1126
Db 1108 TLIAFVSSGCGKSTIQLLERFYDPOGTVMIDGHDKSKVNIQFLASNIGIVSQEPLVD 1167
```


Q9W693
ID Q9W693 PRELIMINARY; PRT; 851 AA.
AC Q9W693;
DT 01-NOV-1999 (TREMREL. 12, Created)
DT 01-NOV-1999 (TREMREL. 12, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE MULTIDRUG RESISTANCE TRANSPORTER HOMOLOGY (FRAGMENT).
GN MDR.
OS Fundulus heteroclitus (Killifish) (Mummichog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomerygia; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Fundulidae; Fundulus.
OX NCBI_TaxID=8078;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER, INTESTINE;
RA Cooper P.S., Van Veld P.A., Reece K.S.;
RT "P-glycoprotein related sequences from the mummichog (Fundulus heteroclitus).";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AF095732; AAD23956.1; -.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC_transporter_tmemb.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane; 1.
DR Pfam: PF00005; ABC_tran; 2.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR ATP-binding; transport.
KW NON_TER 1
FT 1
SQ SEQUENCE 851 AA; 93660 MW; 8F2DC4464171420A CRC64;

Query Match 43.6%; Score 2825; DB 13; Length 851;
Best Local Similarity 63.4%; Pred. No. 3.2e-154;
Matches 541; Conservative 143; Mismatches 153; Indels 16; Gaps 4;

Qy 436 STTVQLMQRLYDPTGMCIDGDIITINVRHLREITGVVSGEPVLEFATTIAENIRYGRE 495
Db 1 STTIQLLREYDPEQGSVSDIRSLNRYLRGMIGVVSQEPILFATTIAENIRYGRP 60

Qy 496 NVTWDETEKAVKANAYDFETKLPNPFDTLVGERGAQLSGGQKQRIATARALVRNPKILL 555
Db 61 DVTETEEKAAKANAYDFETKLPNPFDTLVGERGAQLSGGQKQRIATARALVRNPKILL 120

Qy 556 LDEATSDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGDFDGVIVVEKGNHD 615
Db 121 LDEATSDALDAESETIVQALDKVRQGRITLIVAHCLSTIRNADVIAGLEKGVWELGTHS 180

Qy 616 ELMEKEGIYKLVMTQTRNGNEIELENATGSKSDALEMSPKDSG---SSLIKRRSTRR 672
Db 181 ELMEKGVYHTLVMTQTRNGNEIELENATGSKSDALEMSPKDSG---SSLIKRRSTRR 672

Qy 673 SIHA-----PQOQDRKLGTKEADLNENPVVFWFRILKLNSTWPFVYVGFICAILNGG 725
Db 236 SFFRAASIGERKDKQEKEDKTEDEDFPMVSLFKVLRNASEWPYILVGLICATINGA 295

Qy 726 LQAFSIFPSRIIGTFRDEDETPKRNQSNMFSVLVLGLIISFIITFELQGFEGKAGEI 785
Db 296 IQLPFAVLFSKIITVFA-EPDQTIIRQRANFFSLMFVVGVCVFTFMFGFCFGKSGEV 354

Qy 786 LTKRLRYVFSMLRQDVSFDDPKNTTGALTTRLANDAAQVRKAGISRLAVITQNIANL 845
Db 335 LFLKURLGAFKSMRLQDLGWFDSFKNVSGALTTRELATDAAQVQSGVRLATEAQNIANL 414

Qy 846 GTGIIISLYGQWLTLLALVPIIATAGVVEKMKLSQALKKKKELEGAGKATATAIEN 905
Db 415 GTGVILAFYVWELTLLILAVVPVIALAGAVQMKLTGHAEDKKLEKAGKATATAIEN 474

Qy 906 FTVVYSLRQEKFEYMAQSLQVYRNLSRKAHIFGVFSFISITQAMMYFSYAGCFREGAYL 965
Db 64 GFSVLGIVLGMVTVFLRA---QNSEF---VLGTVSRDPEGLPALTKKEEFDTLVRRYCLY 117

Db 475 IRTVASLTREPKEFESLYEENLVVPYKNSQKHAHYGTFTSFSQAMTYFAYAACFRGAWL 534
Qy 966 VANEFMNQDVLVFSALVFGAMAVGVSSFPADYAKAVSAHVIMIIIEKSPLDISYSP 1025
Db 535 IVEGRMDVEAVFLVISAVLFGAMAVGEANSFAPYAKAKMSASHLMLNKEPEIDNLSE 594
Qy 1026 HGLKPNTLEGNTNEVFNYPTRPDIPVIOGLSLEVKKGTQTLALVSGSGCKSTVQOLL 1085
Db 595 RGEPSDMEDGNVSFEDVKFNYPSPDPVILRLGLNLRVKGKGTALVSGSGCKSTVQOLL 654
Qy 1086 BRFYDPLAGSLVDIGKEIKHLNVOWLRAHLGIYVSOBPIFDCSIAENIAYGDNRSRVVSH 1145
Db 655 BRFYDPRGRVYMDSIDVKRLNIHWRSLQIGIVSQEPLFDCALAEIAYGDNRSRVVME 714
Qy 1146 EIVQAAKEANIHFIETPEKYNTRVGDKGTQSGGQKQRIATARALVRPHILLDEAT 1205
Db 715 EIEAAKAAHNFELPOKYNTQAGDKGTQSGGQKQRIATARALVRPHILLDEAT 774
Qy 1206 SALDTESEKVVQOALDKAREGTCIVTAHRLSTIONADLVFQNGKVKHGTQOQLLAQ 1265
Db 775 SALDTESEKVVQDALDOASKGRTCTIVVAHRLSTIRNADRIAVFGGVVEGQTHQQLLAK 834
Qy 1266 KGIVFSWYSVOAG 1278
Db 835 KGIVHMLVTTQLG 847

RESULT 14
O61301 PRELIMINARY; PRT; 1275 AA.
ID O61301
AC O61301;
DT 01-AUG-1998 (TREMREL. 07, Created)
DT 01-AUG-1998 (TREMREL. 07, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE P-GLYCOPROTEIN.
GN PGP-A.
OS Haemochus contortus
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98226176; PubMed=9566525;
RA Xu M., Molento M., Blackhall W., Ribeiro P., Beech R., Prichard R.;
RT "ivermectin resistance in nematodes may be caused by alteration of P-glycoprotein homolog.";
RL Mol. Biochem. Parasitol. 91:327-335(1998).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AF003908; AAC38987.1; -.
DR HSP: P13569; INBD.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC_transporter_tmemb.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane; 2.
DR Pfam: PF00005; ABC_tran; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; transport.
SQ SEQUENCE 1275 AA; 142204 MW; 7439D3836B75787C CRC64;

Query Match 43.6%; Score 2824.5; DB 5; Length 1275;
Best Local Similarity 44.4%; Pred. No. 6.2e-154;
Matches 568; Conservative 271; Mismatches 416; Indels 25; Gaps 12;

Qy 8 KGSAEKNFWKMGKSKKK---EKKEKKPVSTFAMFRYSNWLDRMLVGMVGTMAALIHGA 63
Db 4 KGQDERIPLIGSKSKSGEVSKEEPPITNRGILSLATTLDYVLLAAGTLPACVHGA 63

Qy 64 ALPLMLVFGNMTDSFANAGISRNKTFPVINESITNNTTOHFINHLEEMTT---YAYY 119
Db 64 GFSVLGIVLGMVTVFLRA---QNSEF---VLGTVSRDPEGLPALTKKEEFDTLVRRYCLY 117

QY 120 YSGICAGVLVAAYIQVSWFLAAGROILKIRKOFFEHAIMRQEIQWFDVHDVDELNRLTD 179
Db 118 YLGLGFAMFATSYIQICWETFAERTHKIRKIYLAAILKQQLSWFDIQOQTGNLTARLFD 177
QY 180 DVSININGIGDKMGFFQSTIATFTGTFVTRGKLTILVILASPVGLGLSAIAWAKILS 239
Db 178 DLERVRGLDGLKSLFTQMSAFAVAGFCVGFAYSWMTLVMMVVAFFIVISANWSKIYA 237
QY 240 SFTDKELLAVAKAAVEAELAAIRTVIAFGGOKKELERYKNLEBAKIGIGIKKAITANI 299
Db 238 TRTOVEQETAVAGAAEETFSIRIVHSGHGRLETRFEAALEKRGQTGLVGYFYMGV 297
QY 300 SIGAAFLLIYASALAFWYGTSLVSS-EYTIQOVLTFFSVLIGAFSGIQAOSPSTEAPA 358
Db 298 GVGFGOMCTVVSVALAFWYGSVLIINDPALDRIGRTFFFAVWMSGSAALGTCLPHLNIIS 357
QY 359 NARGAAVEIPIKIDNPKSIDSYSGHKKPDNTKGNLEFKNVHFSYPSRKEVKILKGLNLK 418
Db 358 IARGAVRSVLSVINSRPKIDPYSLDGIVLNNMRGSIREFKNVHFSYPSRRTLIQLKGVSLQ 417
QY 419 VQSQTVALVGNSSCGCKSTTVQLMORLYDPTDGMVCIDQDITINVRHLREITGVVSOE 478
Db 418 VSAGQKIALVGNSSCGCKSTVNLRLFRYDPTRGKVTIIDDVCDLNVOKLREQIGVVSQE 477
QY 479 PVLFTATTIARNIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGAQLSGQOK 538
Db 478 PVLFDGTLEFNIRKMGYEATMEVQACRVANAADFTKRLPGYTRVGERGVQLSGQOK 537
QY 539 ORTIAIARALVRNPKILLDEATSAIDTSEAEVQVALDKARKGRTTIVIAHRLSTVRNAD 598
Db 538 ORTIAIARALIKPRIILLDEATSAIDTSEAEVQVALDKARKGRTTIVIAHRLSTVRNAD 597
QY 599 VIAGFDGVIVKGNHDKMGKIVYKLVMTQTRNETELENAATGESKESDSEDALEMSPK 658
Db 598 QIFVFKNGTIVEQTHAELMNRGVFEFMTQAQVLRQKEEVLDSDESVDVSPDIALP 657
QY 659 DSGSSILKRSTRSIHA-POQDRLKGTEDLNENVPVPSFWRLIKLNSTEWFFVVGVI 717
Db 658 HSSLSRSKSTRSAISAVPSVRSQI-EMEDLRAKPTPMS-KIFYENRDKWGYIILGL 714
QY 718 FCAIINGGLQPAFSIFRSIIGIFTRDEDPETKRONSNF-SVLFLVLIISFTIFFLQ 776
Db 715 IACITGTVPFAVLYAQIQVIS--EPVDQMKGHVFLWCGAFIVIGLVHAFAPFSA 771
QY 777 FTFGKAGEILLKRLRYMFRSMRDVSWFDDPKNTTGALTTRLANDAAOVKGAIGSKLA 836
Db 772 ICLGRGCEALTKKLFEAFKLLRNQVGYDIDIRHGTGKLCRTATAPNR-VYFTRLP 830
QY 837 VITONANLGTGIIISLIYQWLTLLILAIVPIITAIGAVVEMKMLSGOALKDKKELEGAG 896
Db 831 GVLSSVVTIGALVIGTIFGWLALILMVMVPLIIGSYFEMRMQFGKMRDTELLBEAG 890
QY 897 KIATEAENFRVVSILTREQFEYMAQSLQVPYRNSLRKAHIFGVFSFISTQAMMYSYA 956
Db 891 KVASQAVENIRTVHALNREQOFHEWCEYLKEPYRENLCQHTYGGVAFQSLLFFMYA 950
QY 957 GCFREGAVLVANFNFODVLLVFSIAIVFGAMAVQGVSSFPADYAKAKVSAAHVIMIEK 1016
Db 951 VAFWITGALFVNDHSMQPIIDVYRVFAFMFCQMVGNISSFTPDVVKARLAASLLFYLIEH 1010
QY 1017 SPLIDSYSPHGLKPNLTBGNVTENEVFNVTPTDIPVLOGLSLEVKGKGTALVSSGC 1076
Db 1011 PSEIDNLEDGV-YTKISGHSFRNVFNPTTRQIRVLRGLNLEINPGTFTVALVQSGC 1069
QY 1077 GKSTVVOLLERYDPLAGSVLIDKEIKHLNVQWLRHLGTVSQEPTLFDPCSTAEINAYG 1136
Db 1070 GKSTVMALLERYNQNGVITVDGENIRNNIRNLREQVICVQSEPTLFDCTINENICYG 1129
QY 1137 DNSRVVSHEEIVQAAKEANIHFTETLPEKYNTRVGDKGTQLSGGQKQRTIAIARALVRP 1196
Db 1130 LDDPKPSYEQVAAAKANIHNFVGLPEGYDTRVGEKGTQLSGGQKQRTIAIARALIRP 1189
QY 1197 HILLDEATSAIDTSEAEVQVALDKARKGRTTIVIAHRLSTIQNADLIIVFONGKVKEH 1256

Db 1190 PILLDEATSAIDTSEAEVQVALDKARKGRTTIVIAHRLSTIQDSDVIMVMIQEGKATDR 1249
QY 1257 GTHOOLLAOKGIYFSMVSVQ 1276
Db 1250 GTHEHLLMKNDLYKRICETQ 1269
RESULT 15
Q21349 ID Q21349 PRELIMINARY; PRT; 1321 AA.
AC Q21349;
DT 01-NOV-1996 (TEmBLrel. 01, Created)
DT 01-NOV-1996 (TEmBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TEmBLrel. 19, Last annotation update)
DE K08E7.9 PROTEIN.
GN K08E7.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RP [1]
RP SEQUENCE FROM N.A.
RA Snye R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology".
RL Science 282:2012-2018(1998).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; 277666; CAB01232.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmam.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP-GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1321 AA; 145062 MW; 68094121B30E8746 CRC64;
Query Match 42.4%; Score 2743; DB 5; Length 1321;
Best Local Similarity 45.1%; Pred. No. 3,2e-149;
Matches 585; Conservative 219; Mismatches 435; Indels 58; Gaps 12;
QY 21 KSKKKEKKKPTVSTFAFRYSNWLDRLYMLVGTMAIITHGAALPLMLVFGNMTDSPA 80
Db 47 KITRAKEEVNKNVSIPOLYRITTTLEKLLFGLTAVITGAGLPLMSILQKVSQAFI 106
QY 81 NAGISRNKTFPVIINESTNNTQHFI-----NHLEEMTTYYAYISGIGAGVLVAAY 132
Db 107 NE-----QIVN-----NNGSTFLPTGQNTKTFEHDVNMVWVSYAAMTVGMWAAGQ 154
QY 133 IQVSWFLAAGROILKIRKQFFHAIMRQEIQWFDVHDVDELNRLTDVDSKINEGIGDKI 192
Db 135 ITVTCLYVAEOMNRLREFEVKSILRQETSWEDTNSGTLATKFLDNLRVKEGTGDKI 214
QY 193 GMFFOSIAFTFTGTVGRGKLTILVILASPVGLGLSAIAWAKILSSFTDKELLAYAKA 252
Db 215 GMAFYLSQFTIGTGFVAFTHSWQTLVVALVTPIQALCGFAIAKSMSTFAIRETLRYAKA 274
QY 253 GAVAEVLAARTVAFGGOKKELERYKNLEAKGIGIKKAITANISIGAAFLLIYASY 312
Db 275 GKVVEETISSRTVSVNLGRYLELERYSTAVEEAKKAGVLKGLFLGIFSGGAQSNFISF 334
QY 313 ALAFWYGTSLVLSSEYTIQOVLTFFSVLIGAFSGIQAOSPSTEAFANARGAAVEIPIKID 372
Db 335 ALAFYIGVGVHDSGLNFGDMLTFTSSVMGSMALGLAGPQLAVLGTAGGAASGIYEVD 394

QY 373 NKPSIDSYKSGHKPDNIKGNLEKFNHVSYPKSRKEVKILKGLNLKVOSGOTVALVGNSSG 432
Db 395 RRPVIDSSSKAGKDKMGIDVTNVEHFTTPSRPDVPIILRGMLRVNAGOTVALVGSNG 454
QY 433 CGKSTTVQMLQRLYDPTDGMVCGIDGDIINTINVRHLREITGVVQSEPVLFATTIAENIRY 492
Db 455 CGKSTIISLLRYDYDLKGIIDGVDRDINLEFLRKNVAVVQEPALFNCTIENISL 514
QY 493 GRENVTMBIEKAVKEANNAYDFIMKLPKFDTLVGERGAQLSGGOKORATARALVRNPK 552
Db 515 GKEGITREEMVAACKMANAEKIKTPNGYNTLVDRGTQLSGGOKORATARALVRNPK 574
QY 553 ILLDEATSALDESEFVQVQALDKARGRITVIAHRLSTVRNADVIAGDDGVIVEKG 612
Db 575 ILLDEATSALDAESGIVQOALDKARGRITVIAHRLSTVRNADVIAGDDGVIVEKG 634
QY 613 NHDELKMKGIYFKLVMTMOTRNEIE-----LENATGESKESDALEMSPKDSGSSL 664
Db 635 DHRALMAQOGLYDLVTAQFTDAVDSAAEGKFSRENSVARQTSHEGLSRQASEMDDIM 694
QY 665 IK-RRSTRSI-HAQO-----GQDRKLGTCKEDLNENVPVFWIRILKLNSTEMPY-- 712
Db 695 NRVRSSTIGSITNGPVIDEKEERIGKDALSRILKQLEEN-----NAQTNLFEILYHA 747
QY 713 -----FVVGIFCAIINGLOPAFSIIFSKIIIGIFTRDEDPETKRONSMFSLFLVLGI 766
Db 748 RPHALSIFGOMSTATIGGYIPTYFTSPMNVFA--GNPADFLSQGHFWMFLVLAA 805
QY 767 ISFITFLQAGTFEGKAGEILTKRLYRVFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQ 826
Db 806 AQGICSFMTFFMGIASESILTRDLRKNLFRNLVSHIGFFSDPQNASGKISTRLATDPEN 865
QY 827 VKGAIGSRNAVITQNTANLGTGIIISLYGWLTLALLAIVPIIATAGVVMKMLSGOAL 886
Db 866 LRTAIDFRFSTVITTLVSMVAGIGLAFYQWQWALLIIAILPIVAFGQYLRGRFTGKNV 925
QY 887 KDKKELEGAGKIATEIENRVTVSLTREQFEYMYAQSLQVRYNLSRAHIFGVSPSI 946
Db 926 KSAEFADSKIAEIAENVRTVOALAREDTFYENFCEKLDIPHKEAIEAFIQGLSYG 985
QY 947 TQAMMYFSYACFRFGAYLVANEFMFOD--VLLVFSVIFGAMAVGVSSFAPDYAKAK 1004
Db 986 ASSVLYLLNTCAVRMGLALITDPTMQPMRVLRVMAITISTLTGFATSFPEYAKAT 1045
QY 1005 VSAAHVIMIEKSLPIDSYPHGLKPNLTLEGNTFNEVFNPTRPDIPVLQGLSLEVKK 1064
Db 1046 FAGIIFGMLRKISKIDSLAGEK-KKLYGKVFKNVRFAYPERPEIEILKGLSPSVEP 1104
QY 1065 GQTLALVSSGCGKSTVQVQLLERYDPDLGKVLIDGKEIKHLNVQWLRHLGIYSQEPIL 1124
Db 1105 GQTLALVSPGCGKSTVALLERFYDPLGGEIFIDGSEIKTLNPEHTRSQIAIVSQEPTL 1164
QY 1125 FDCSIAENIAGDSRVVSHEEIVQAAKEANIIHFIETLPEKYNTRVGDGTOLSGGOK 1184
Db 1165 FDCSIAENIYGLDPSVTAQVEEAARLANIHNFIHELPEGETRVGDGTOLSGGOK 1224
QY 1185 RIAIARALVRPHILLDEATSALDESEKVVQVQALDKAREGRITCIVIAHRLSTIONADL 1244
Db 1225 RIAIARALVRNPKILLDEATSALDESEKVVQVQALDKAREGRITCIVIAHRLNTVMNAD 1284
QY 1245 IVFQNGKVKHGHQQLLAQKGIYFSVSVQAGAKR 1281
Db 1285 IAVVSNGTIIEKGTHTQLMSEKGAAYKLTOKQMTTEK 1321

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 6, 2002, 19:26:40 ; Search time 2230.96 Seconds
(without alignments)
7749.846 Million cell updates/sec

Title: US-09-672-725C-27

Perfect score: 6473

Sequence: 1 MDPGGRKGSAGNEKFWKMGK.....LLAQKGIYFSVMVQAGAKR 1281

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgr2_1/USPTO.spool/US09672725/runat_04112002_124342_18001/app_query.fasta_1.6435
-DB=EST -OPMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -List=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -NODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725.cgrn_1_17365@runat_04112002_124342_18001 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEVTIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1424	22.0	943	9	AL520322
2	1073.5	16.6	981	10	BF969667

3	1035	16.0	894	10	BF584568
4	1031	15.9	803	10	BG298756
5	1027	15.9	1019	10	BG248052
6	971.5	15.0	944	10	BF796582
7	958	14.8	851	10	BF969062
8	953	14.7	785	9	AV709991
9	919.5	14.2	795	10	BF313560
10	919	14.2	726	10	BG293345
11	911.5	14.1	628	10	BG080311
12	910	14.1	587	10	BM053000
13	889	13.7	563	10	BI964198
14	889	13.7	563	10	BI964218
15	889	13.7	563	10	BM052878
16	880	13.6	605	9	AV616675
17	875	13.5	563	10	BM053038
18	875	13.5	2676	11	AK014319
19	826	12.8	546	10	BI964137
20	824.5	12.7	1201	10	BF304387
21	808	12.5	759	10	B8376459
22	806.5	12.5	871	12	A2682350
23	792	12.2	662	10	BG963638
24	791.5	12.2	886	12	BH139685
25	780	12.1	517	10	BI964115
26	778.5	12.0	897	12	A2541090
27	772	11.9	780	10	BG587938
28	768.5	11.9	932	12	A2670821
29	768	11.9	560	10	BE749379
30	768	11.9	822	10	BI332761
31	764.5	11.8	834	12	A2548312
32	764.5	11.8	947	12	A2683753
33	763.5	11.8	913	12	BH155700
34	760.5	11.7	880	12	A2687805
35	760	11.7	669	10	BM013981
36	752.5	11.6	823	12	A2532602
37	750.5	11.6	853	12	A2679807
38	748.5	11.6	919	12	A2690701
39	747.5	11.5	511	9	AA239727
40	745.5	11.5	886	12	A2540627
41	745	11.5	699	9	BB667773
42	743	11.5	600	10	BM486593
43	741.5	11.5	891	12	A2682250
44	734	11.3	760	10	BM016204
45	732.5	11.3	899	12	BH154857

ALIGNMENTS

RESULT 1	AL520322	LTI_NFL004_NBC2	943 bp	mRNA	linear	EST 13-FEB-2001
AL520322	Prime, mRNA sequence.					CS0DB006YC15 5
DEFINITION	AL520322	LTI_NFL004_NBC2	Homo sapiens	cDNA clone		
ACCESSION	AL520322	EST.				
VERSION	AL520322.1	GI:12783815				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
TITLE	1 (bases 1 to 943)					
JOURNAL	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.					
COMMENT	Full-length cDNA libraries and normalization					
	Unpublished (2001)					
	Contact: Genoscope					
	Genoscope - Centre National de Sequencage					
	BP 191 91006 EVRY cedex - France					
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.					
FEATURES	Location/Qualifiers					
Source	1. 943					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="CS0DB006YC15"					
	/clone_lib="LTI_NFL004_NBC2"					

QY	1217	GinGluaLeuAaspLySaLaAArgGlUGlyArgThrCysIlleValIleAlaHisArgLeu	1236
Db	722	CAAGAAGCCCTGGACAAAGCCAGAGAAGCGCGACCTGCATTGTGATGTCTACCGCGTG	781
QY	1237	SerThrIleGlnAsnAlaSplLeuIleVal-ValPheGlnAsnGlyLysValLysGluHis	1256
Db	782	TCACCATTCAGATGCAGACTTATTAGTGTGTGTTTCAGAAATGCGCAGAGTCAAGGAGCA	841
QY	1256	sGlyThrHisGinGlnLeuLeuAlaGlnLysGlyIleIlyrPheSerMetValSerValG	1276
Db	842	TGGCAGCGATCAGCAGCTCTGGCAGAAAGGCATCTATTTTCAATGGTTCAGTGTC	901

	Qy	1276	nAlaGlyAlaLysArg	1281
		000	GGGGGGGGGGGGGG	017

RESULT 2	BF969667	BF969667	981 bp	linear	EST 22-JAN-
LOCUS	602272046f1	NTH_MGC_84	Homo sapiens	cdna clone	IMAGE:4360090 5'
DEFINITION	mRNA sequence.				
ACCESSION	BF969667	1	CT.12235992		

KEYWORDS EST.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.

Email: cgap@ps-femail.nlm.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The T M A G E Consortium (LMNL)

DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

```

FEATURES
  source
    Location/Qualifiers
      High quality sequence stop: 670.

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:436090"
/clone_lib="NIH MGC 84"

```

```

/tissue_type="adrenal cortex carcinoma, cell line
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site:
united states; Cell: cloned unidirectionally; Oligo-dT

```

primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COOK 207 3 151 3
ORIGIN
Alignment Scores:

Pred. No.:	4,135-62	Length:	234
Score:	1073.50	Matches:	234
Percent Similarity:	86.60%	Conservative:	18
Best local similarity:	80.41%	Mismatches:	31

Query Match:	16.58%	Inders:	10
DB:	10	Gaps:	5

Qy 793 MetValPheArgSerMetLeuArgGlnAspValSerTrpPheAspPProLysAsnThr 812

DOI 10.1002/anie.201100081

BASE COUNT	267 a	194 c	250 g	270 t
ORIGIN				
Alignment Scores:				
Pred. No.:	4.15e-82			Length: 981
Score:	1073.50			Matches: 234
Percent Similarity:	86.60%			Conservative: 18
Best Local Similarity:	80.41%			Mismatches: 31
Query Match:	16.58%			Indels: 10
DB:	10			Gaps: 5
US-09-672-725C-27 (1-1281) X BF969667 (1-981)				

Qy 793 MetValPheArgSerMetLeuArgGlnAspValSerTrpPheAspProLysAsnThr 812

+ 1 aacccmnmccccamcacaatcatcacacaccagcatgcactgctgttcgatgcccttaaacacc 60

Db	541	ATCATGAAACTGCCCAACATTTGACACCTCTGGTGGTGGAGAGGGGGCCGACGTGAGT	600
QY	535	GlyGlyGlnGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu	554
Db	601	GGGGGACAGAAACAGAGAATCCG-ATTGCCCGGGCCCTGGTCCCGCAATCCCAAGTGCCTT	659
QY	555	LeuLeuAspGlnAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla	574
Db	660	TC-GTGACGAAGCCACCTCAGC-CTGGATACAGAAAGTGAAGCTGGGGTTTCAGGCGCGCA	717
QY	575	-LeuAspLysAlaArgLysGlyArgThrIle-----ValIleAlaHisArg	590
Db	718	CCTAGATAGGGTACAAAGAGGGGNGGACACATTGGGAAGCTCATCGCTGTACCGTCGAA	777
QY	590	gLeuSerThrValArg 595	
Db	778	GCTGACGGCCATCGG 793	
RESULT 4			
LOCUS	602396681F1 NIH_MGC_94	803 bp mRNA linear EST 21-FEB-2001	
DEFINITION		Mus musculus cDNA clone IMAGE:4511157 5',	
ACCESSION	BC298756	mRNA sequence.	
VERSION	BC298756.1	GI:13063728	
KEYWORDS		EST.	
SOURCE		house mouse.	
ORGANISM		Mus musculus	
REFERENCE			
AUTHORS		1 (bases 1 to 803)	
TITLE		NIH-MGC http://mgi.nci.nih.gov/	
JOURNAL		National Institutes of Health, Mammalian Gene Collection (MGC)	
COMMENT		Unpublished (1999)	
		Contact: Robert Strausberg, Ph.D.	
		Email: cgabbs-remail.nih.gov	
		Tissue Procurement: The Cepko Laboratory	
		cDNA Library Preparation: Life Technologies, Inc.	
		CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	
		CDNA Sequencing by: Incyte Genomics, Inc.	
		Clone distribution: MGC clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LLNL at:	
		http://image.llnl.gov	
		Plate: LLAM10394 row: b column: 22	
		High quality sequence stop: 698.	
FEATURES			
source		Location/Qualifiers	
		1..803	
		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		/clone="IMAGE:4511157"	
		/clone_lib="NIH_MGC_94"	
		/tissue_type="retina"	
		/lab_host="DH10B (phage-resistant)"	
		/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NOTI;	
		Site_2: SalI; Cloned unidirectionally; oligo-dT primed.	
		Average insert size 3.3 kb. Library enriched for	
		full-length clones and constructed by Life Technologies.	
		Note: this is a NIH_MGC Library."	
BASE COUNT		240 a 162 c 205 g 196 t	
ORIGIN			
Alignment Scores:			
Pred. No.:	1,478-78	Length:	803
Score:	1031.00	Matches:	222
Percent Similarity:	91.51%	Conservative:	15
Best Local Similarity:	85.71%	Mismatches:	21
Query Match:	15.93%	Indels:	4
DB:	10	Gaps:	0
US-09-672-725C-27 (1-1281) x BG298756 (1-803)			
QY	236	IsvtIleLeuSerPheThrAspLysGluLeuAlaThrValArgAlaValAlaVal	255

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: L1A10334 row: n column: 21
High quality sequence stop: 650.

FEATURES

Location/Qualifiers
1..1019
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:488404"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 276 a 262 c 283 g 196 t 2 others
ORIGIN

Alignment Scores:

Pred. No.: 4,31e-78 Length: 1019
Score: 1027.00 Matches: 217
Percent Similarity: 90.48% Conservative: 11
Best Local Similarity: 86.11% Mismatches: 19
Query Match: 15.87% Indels: 5
DB: 10 Gaps: 2

US-09-672-725c-27 (1-1281) x BG248052 (1-1019)

```
QY 1033 LeuGluGlyAsnValThrPheAsnGluValPheAsnTyrProThrArgProAspIle 1052
Db 1 TTGAAGGAATGTCGAATTTAGTGAGCGTCGTTCACTATCCACCACCGCCAGCATC 60
QY 1053 ProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGly 1072
Db 61 CCAGTGTCTCAGGGGTGAGCCTTGTAGGTGAAGAGGGCCAGACGCTGGCCCTGGTGGC 120
QY 1073 SerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeu 1092
Db 121 AGCAGTGGCTGCGGGAAGACACAGTGGTCCAGCTGCTCGAGCGCTTCTACGACCCCATG 180
QY 1093 AlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArg 1112
Db 181 GCTGGATCAGTGTTCATAGTCGCAAGAAATAAAGCAACTGAATGTCATGCTGCCGCA 240
QY 1113 AlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsn 1132
Db 241 GCACAGCTGGGATTTGTGTCCCAAGAGCCCATCTCTTTGACTGCAGCATCGCAGAGAAC 300
QY 1133 IleAlaTyrGlyAspAsnSerArgValValSerHisGluGluIleValGlnAlaAlaLys 1152
Db 301 ATTGCTACGAGACACACAGCGGGTGTCTTATGAGGAGATTGTGAGGCGACGCCAAG 360
QY 1153 GluAlaAsnIleHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGly 1172
Db 361 GAGGCCAACATCCACAGTTCATCGACTCGCTACCTGATAATACACACAGAGTAGGA 420
QY 1173 AspLysGlyThrGlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeu 1192
Db 421 GACAAAGGCACCTACGCTCGGGTGGGAGAGCAGCGCATCGCCATCGCAGCGGCCCTC 480
QY 1193 ValArgGlnProHisIleLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer 1212
Db 481 GTCACAGACCTCACATTTTACTTCGACCAAGCAACATCAGCTCTGGATACAGAAGT 540
QY 1213 GluLysValValGlnGlnAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIle 1232
Db 541 GAAAGGTTGTCAGGAAGCGCTGCACAAACCCAGGAGGCCGCCACCTGCATGTGTATC 600
QY 1233 AlaHisArgLeuSerThrIleGlnAsnAlaAspLeuIle-ValVal-PheGlnAsnGlyL 1252
```

```
Db 601 GCTCACCGCTGTCACCATCCAGAACGGGACTTGATGGTGGTATTCAAGAACCGGA 660
QY 1252 ysValLys--GluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheS 1271
Db 661 AGGTCAAGGAGGACGCGGACCCCAACAGCAAGTGTGGCGCAGAGGCACTACTTCT 720
QY 1271 erMetValSer-ValGlnAlaGlyAlaLys 1280
Db 721 CCAGG--TCAGGTGCATGCTGGCGCAAG 747
RESULT 6
BF796582
LOCUS BF796582 944 bp mRNA linear EST 12-JAN-2001
DEFINITION 6022584G3F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:434170 5',
mRNA sequence.
ACCESSION BF796582
VERSION BF796582.1 GI:12101636
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue procurement: Louis Staudt, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9955 row: n column: 15
High quality sequence start: 9
High quality sequence stop: 669.
```

FEATURES

Location/Qualifiers
1..944
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:434170"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 265 a 241 c 237 g 201 t
ORIGIN

Alignment Scores:

Pred. No.: 2,32e-73 Length: 944
Score: 971.50 Matches: 211
Percent Similarity: 83.64% Conservative: 19
Best Local Similarity: 76.73% Mismatches: 39
Query Match: 15.01% Indels: 8
DB: 10 Gaps: 1

US-09-672-725c-27 (1-1281) x BF796582 (1-944)

```
QY 988 MetaLaValGlyGlnValSerSerPheAlaPro-Asp-TyrAlaLysAlaLysValSerA 1007
Db 1 GTGGCTCTAGGACATGCCAGTTCATTTGCTCCAGGAGGTATGCTAAAGCTAAAGCTGCTG 60
QY 1007 laAlaHisValIleMetIleIleGluLysSerProLeuIleAspSerTyrSerProHisG 1027
Db 61 CAGGCCACTTATTTCATGCTGTTTGAAGAACAAACCTCTGATTGACAGCTACAGTGAAGAGG 120
```

```
QY 1027 lyleuLysProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrP 1047
|||||
Db 121 GGCTGAAGCCTGATAAATTTGAAGGAATATACATTTATGAAGTGTGTTCAACTATC 180
|||||
QY 1047 rothrArgProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnT 1067
|||||
Db 181 CCACCCGAGCAACAGTCGACGCTTCAGGGCTGAGCTGAGGTGAAGAAAGCCAGA 240
|||||
QY 1067 hrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluA 1087
|||||
Db 241 CACTAGCCCTGGTGGCAGCAGTGGCTGTGGAAGACGACGCTGTCACAGCTCCTGGAGC 300
|||||
QY 1087 rgPheTyrAspProLeuAlaGlySerValLeuLeuAspGlyLysGluIleLysHisLeuA 1107
|||||
Db 301 GGTCTACGACCCCTTGGCGGGACAGTCTCTCATGCTGCAAGAACGAAGAACTCA 360
|||||
QY 1107 snValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspC 1127
|||||
Db 361 ATCTCCAGTGGCTCAGAGCTCAACTCGGAATCGTCTCAGGAGCCTATCTCTATTGACT 420
|||||
QY 1127 ysserIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgValValSerHisGluGluI 1147
|||||
Db 421 GCAGCATTCGCGAGAATATTGCCCTATGGACACACAGCCGGTGTGTATCACAGGATGAAA 480
|||||
QY 1147 leValGlnAlaAlaLysGluAlaAsnIleHisPheIleGluThrLeuProGluLysT 1167
|||||
Db 481 TCGTGAGTCAGCCAAAGCTGCCAATACATCTCTTCATCGAGAGCTTACCCCAAAAT 540
|||||
QY 1167 yrAsnThrArgValGlyAspGlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleA 1187
|||||
Db 541 ATGAACAAGAGTGGGAGATAAGGGACTCAGCTCTCAGGAGGTCACAAACAGAGATTG 600
|||||
QY 1187 laIleAlaArgAlaLeuValArgGln-ProHisIleLeuLeuLeuAspGluAlaThrSer 1206
|||||
Db 601 CTATGTCGCGAGCCCTCATCAGACACACCCCTCAAAATCTCTCTGTGTGATGAGTACATCA 660
|||||
QY 1207 Ala-LeuAspThr-----GluSerGluLysValValGlnGluAlaLeuAspLysAlaR 1224
|||||
Db 661 GCCTCTGGATACCTGACAAAGTGAACAGGTTGTCCCAAGAAAGCCCTGTGCACACAGAGA 720
|||||
QY 1224 gGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAlaAspLe 1244
|||||
Db 721 AGAGCGCGCACCTGCTGTGTGATGTCACCGCTGTC--AACATCCGAAGTGCAGACTT 778
|||||
QY 1244 uIleValValPheGlnAsnGlnLysValLysValLysGluHis 1256
|||||
Db 779 -ATAGTGTGTCTCCCAAGCGGAGAGTCAAGACACAT 814
|||||

RESULT 7
BF969062 851 bp mRNA linear EST 22-JAN-2001
LOCUS 60226993F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4358135 5',
DEFINITION mRNA sequence.
ACCESSION BF969062
VERSION BF969062.1 GI:12336277
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 851)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

Plate: LLAM9995 row: j column: 24
High quality sequence stop: 612.

FEATURES
source

1. .851
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4358135"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="pH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site: 1;
NotI; Site: 2; SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 243 a 172 c 214 g 222 t

ORIGIN

Alignment Scores:

Pred. NO.: 2.97e-72 Length: 851
Score: 958.00 Matches: 217
Percent Similarity: 79.25% Conservative: 16
Best Local Similarity: 73.81% Mismatches: 33
Query Match: 14.80% Indels: 29
DB: 10 Gaps: 5

US-09-672-725c-27 (1-1281) x BF969062 (1-851)

QY 826 GlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeu 845
|||||
Db 3 CAAGTTAAAGGGCTATAGGTTCCAGGCTTGTGTATATACCAGATATAGCAATCTT 62
|||||
QY 846 GlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAla 865
|||||
Db 63 GGGCAGCAAGTAATATATCTCTATGTTGGCAACTAACACTGTTACTTTAGCA 122
|||||
QY 866 IleValProIleIleAlaIleAlaGlyValValGluMetLysMetLysGlyGlnAla 885
|||||
Db 123 ATTGTACCCATCATTCGAATAGCAGAGTGTGTGAATGAAGAAATGTTCTGGACAGCA 182
|||||
QY 886 LeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsn 905
|||||
Db 183 CTGAAGATAAGAAAGAACTAGAGGTGCTGGGAAGATCGTACTGAAGCAATAGAAAAC 242
|||||
QY 906 PheArgThrValValSerLeuThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSer 925
|||||
Db 243 TTCCGAACCGTGTCTTCTTGTACTCAGGAGCAAGTGTGAACATATATGTCTCAGAGT 302
|||||
QY 926 LeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGlyValSerPheSer 945
|||||
Db 303 TTGCAGGTACCATACAGAAACCTTTGAGGAAGGCACACATCTTTGGAATTACATTTCC 362
|||||
QY 946 IleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeu 965
|||||
Db 363 TTCACCCAGCAATGATGATTTTCTCTATGCTGGATGTTCCGGTTTGGAGCCCTACTTG 422
|||||
QY 966 ValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSerAlaIleValPhe 985
|||||
Db 423 GTGGCACAATAAATCATAGCTTGTGAGGATGTTCTGTAGTATTTTACCTGTTGCTTT 482
|||||
QY 986 GlyAlaMet-AlaValGlyGlnValSerPheAlaProAspTyrAlaLysAlaLysVa 1005
|||||
Db 483 GGTGCCATGGCGGTGGGCAAGTCAGTTCAITGCTCTGACTATGCCAAAGCCAAAT 542
|||||
QY 1005 lserAlaAlaHisValIleMetIleGluLysSerProLeuIleAspSerTyrSerPr 1025
|||||
Db 543 ATCAGCAGCCCATCATCATGATCATTTGAAAAAACCCTTTGAT-GACAGCTACAGCAC 601
|||||
QY 1025 oHisGlyLeuLysProAsnThr--LeuGluGlyAsnValThrPhe-----AsnGluVa 1042
|||||
Db 602 GGAAGGCCATAATTCGCGGAACACATTTGGGAAGAAATGTCACCATCTGGTGAACGTTGT 661
|||||
QY 1042 lValPheAsnTyrProThrArg-ProAspIleProValLeuGlnGlyLeuSerLeuGluV 1062


```

/clone="IMAGE:4129222"
/clone_lib="NIH_MGC_19"
/lab_host="DHI0B (phage-resistant)"
/notes="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT      213 a 189 c 236 g 157 t
ORIGIN
Alignment Scores:
Pred. No.:      5,57e-69      Length:      795
Score:          919.50      Matches:    211
Percent Similarity: 87.50%      Conservative: 13
Best Local Similarity: 82.42%      Mismatches: 29
Query Match:    14.21%      Indels:     9
DB:             10      Gaps:      2
US-09-672-725C-27 (1-1281) x BF313560 (1-795)
QY 1017 SerProLeuIleAspSerTySerProHisGlyLeuLysProAsnThrLeuGluGlyAsn 1036
Db 2 AACCCCTTTGATGACAGCTACACGACGAGGAGCCCTAATGCCGACACATG-CAAGGAAT 60
QY 1037 ValThrPheAsnGluValValPheAsnTyProThrArgProAspIleProValLeuGln 1056
Db 61 GTACATT-GGTGAGT-GTATTCACTATCCACCCGACCGGACATCCAGTGTTCAG 118
QY 1057 GlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCys 1076
Db 119 GGACTGAGCCTGGAGGTGAAGAGGCGACAGCTGGCTCTGGTGGCGACAGTGGCTGT 178
QY 1077 GlyLysSerThrValValGlnLeuGluArgPheTyAspProLeuAlaGlySerVal 1096
Db 179 GGGGAAGACACAGTGGTCCAGCTCCCTGGAGCGGTTCACGACCCCTTGGCAGGGAAGTG 238
QY 1097 LeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGly 1116
Db 239 CTGCTTGATGCAAGAAATAAAGCGACTGAATGTTTCAGTGGCTCCGAGCACACCTGGGC 298
QY 1117 IleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyGly 1136
Db 299 ATCGTGTCCAGGAGCCCATCCTGTTTGACTGCAGCATTCCTGAGACATTCGCTATGGA 358
QY 1137 AspAsnSerArgValValSerHisGluGluIleValGlnAlaLysGluAlaAsnIle 1156
Db 359 GACAACAGCCGGTGGTGTACAGGAAGAGATCGTGAGGGCAGCAAGAGGAGGCAACATA 418
QY 1157 HisHisPheIleGluThrLeuProGluLysTyAsnThrArgValGlyAspLysGlyThr 1176
Db 419 CATGCTTCATCGACTCACTGCCTTAATAATATAGCACTAAAGTAGAGAGCAAGGAAC 478
QY 1177 GlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnPro 1196
Db 479 CAGCTCTCTGGTGGCCGAAACAACAGCATTCGCCATAGCTCGTGCCTTGTAGACAGCCT 538
QY 1197 HisIleLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysVal-Va 1216
Db 539 CATATTT-GCTTTGGTAGAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTGCCT 596
QY 1216 IclnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLe 1236
Db 597 CCAAGAAGCCCTGGCAAGACGACAGAGAGG- --GACCTGCATGTGATGTCTACCCCTG 653
QY 1236 uSerThrIleGlnAsnAlaAspLeuIleValIleValPheGlnAsnGlyLysValysGluHI 1256
Db 654 -TCCACCATCCAGATGAGA--CTATAGTGGTGTTCAGAAATGCCAGAGTCCAGGGGGC 710

```

```

QY 1256 sGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyPheSer 1271
Db 711 ATGGACCATCAGAGGTGCTGGACAGAAAGG--CTCTATTTTCA 753
RESULT 10
BG293345 726 bp mRNA linear EST 21-FEB-2001
LOCUS 602390738F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4502552 5',
DEFINITION mRNA sequence.
ACCESSION BG293345
VERSION BG293345.1 GI:13052943
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 726)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Prepared: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10371 row: 1 column: 09
High quality sequence stop: 658.
Location/Qualifiers
1..726
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4502552"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DHI0B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 189 a 220 c 190 g 127 t
ORIGIN
Alignment Scores:
Pred. No.:      5,52e-69      Length:      726
Score:          919.00      Matches:    186
Percent Similarity: 94.66%      Conservative: 9
Best Local Similarity: 90.29%      Mismatches: 2
Query Match:    14.20%      Indels:     2
DB:             10      Gaps:      0
US-09-672-725C-27 (1-1281) x BG293345 (1-726)
QY 1068 LeuAlaLeuValGlySerSerGlyCysGlyLysSerThrValGlnLeuGluArg 1087
Db 2 CTGGCCCTGGTGGCAGCAGTGGCTCGGGAAGAGACAGTGGTCCAGCTGCTCAGAGCGC 61
QY 1088 PheTyAspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsn 1107
Db 62 TTCTATGACCCCATGGCTGGATCAGTGCCTTTAGATGGTCAAGAGCAAGAACTCAAT 121
QY 1108 ValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCys 1127
Db 122 GTCCAGTGGCTCCGAGCTCAACTGGGCAATTTGTCTCCAGAAACCCCATCTCTTTGACTGC 181
QY 1128 SerIleAlaGluAsnIleAlaTyGlyAspAsnSerArgValValSerHisGluGluIle 1147
Db 182 AGCATCGCAGAGAAACATCGCCTATGGAGACACAGCCGGGTGCTGCCTCATGATGAGATT 241

```

FEATURES

source

QY 1148 ValGlnAlaLysGluAlaAsnIleHisHisPheIleGluThrLeuProGluLysTyr 1167
Db 242 GTGAGGCGACCCAGGAGGCGACATCCACCCCTTCATCGACAGCTGCCCAAAATAT 301
QY 1168 AsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGlyGlnArgIleAla 1187
Db 302 AACACAGAGTAGGAGACAGGAGGAGCGAGCTCTGTGGGGCCAGAGGAGGATTGCC 361
QY 1188 IleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAla 1207
Db 362 ATGCCCGAGCCCTCATCAGACAGCTCGGTCCTACTGCTGGATGAAGCCACGTCACGT 421
QY 1208 LeuAspThrGluSerGluLysValValGlnGluAlaLeuAspLysAlaArgGluGlyArg 1227
Db 422 CTGGATCTAGAGAGTCAAAAGGTTGTCCAGGAAGCACTGGCAACAGCCAGGAGGCGGC 481
QY 1228 ThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAla-AspLeuIleValVa 1247
Db 482 ACCTGATTCGATCGCTCACCAGCTGTCCACCATCCAGAGCGCGCACCTTGATCGTGGT 541
QY 1247 lPheGlnAsnGlyLysValLysGluHisGlyThr-HisGlnGlnLeuLeuAlaGlnLysG 1267
Db 542 GATTGAGAGCGGCAAGGTCAAGGAGCGAGCGACCCACCCAGCAGAGCTGCTGGCGGAGAGG 601
QY 1267 lyleTyrPheSer 1271
Db 602 GCATCTATTCTCA 615

RESULT 11
LOCUS BG080311 628 bp mRNA linear EST 26-JAN-2001
DEFINITION H3052B06-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3052B06 5', mRNA sequence.
ACCESSION BG080311
VERSION BG080311.1 GI:12562879
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 628)
AUTHORS Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
T.S., Carter,M.G. and Ko,M.S.H.
Verification and initial annotation of NIA mouse 15K cDNA clone set
Unpublished (2001)
Other ESTs: H3052B06-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.
plate: H3052 row: B column: 06
Seq primer: -21M13 Reverse
High quality sequence stop: 628
POLYA-No.

FEATURES
source Location/Qualifiers
1..628
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niaEST:H3052B06-5"
/db_xref="taxon:10090"
/clone="H3052B06"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic

part of E7.5 embryos, extraembryonic part of E7.5 embryos , and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 157 a 168 c 181 g 122 t
ORIGIN

Alignment Scores:
Pred. No.: 2,05e-68 Length: 628
Score: 911.50 Matches: 187
Percent Similarity: 75.58% Conservative: 8
Best Local Similarity: 72.48% Mismatches: 12
Query Match: 14.08% Indels: 51
DB: 10 Gaps: 2

US-09-672-725C-27 (1-1281) x BG080311 (1-628)

QY 1024 SerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGluValVal 1043
Db 1 AGCACAGAGGCGTGAAGCGCTACTCTGTAGAGGAATGTAAATTTATGGAGTCCAG 60
QY 1044 PheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeuGluValLys 1063
Db 61 TTTAATCTCCACCCGACCAACATCCAGTGTCTTACGGGGCTGAGCCTCAGGTGAAG 120
QY 1064 LysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrValValGln 1083
Db 121 AAGGGCCAGACGTTGGCCCTGTGGGGCAGCAGTGGCTGTGGAGAGACACAGTGGTCCAG 180
QY 1084 LeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGlyLysGluIle 1103
Db 181 CTGCTCGAGCGCTTCTACGACCCCATGGCTGATCAGTGTCTTAGATGGCAAGAAATA 240
QY 1104 LysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSerGlnGluProIle 1123
Db 241 AAGCAACTGAATGTCCAGTGGCTCCGAGCTCACCTTGGCATTTGTGCCAGGAGCCATT 300
QY 1124 LeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgValValSer 1143
Db 301 CTCCTTGTACTGCAGCATTCGAGAGAACATCGCCTATGGAGACAACAGCCGGCGGTGCT 360
QY 1144 HisGluGluIleValGlnAlaAlaLysGluAlaAsnIleHisHisPheIleGluThrLeu 1163
Db 361 CATGAGGAGATTGTGAGGGCAGCCAAAGGAGGCCAACATCCACCATTCATCGACTCACTG 420
QY 1164 ProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGlnLys 1183
Db 421 CCTGAT-----
QY 1184 GlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeuAspGlu 1203
Db 426 -----
QY 1204 AlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeuAspLysAla 1223
Db 427 -----GTTCTCCAGGAGCGCTGGACAAGCC 453
QY 1224 ArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAlaAsp 1243
Db 454 AGGGAAGGCGCCACCTGATTTGTATCGTCACCGCGCTGTCCACCATCCAGAGACGGGAC 513
QY 1244 LeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGlnGlnLeuLeu 1263

[illegible]

Score: 889.00 Matches: 171
Percent Similarity: 96.26% Conservative: 9
Best Local Similarity: 91.44% Mismatches: 7
Query Match: 13.73% Indels: 0
DB: 10 Gaps: 0

US-09-672-725C-27 (1-1281) x BI964218 (1-563)

QY	950	MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu	969
DB	3	ATGATGATATTTTCCCTATGCTGGATGTTTCCGGTTTGGAGCCTACTTGGTGGCACAATAA	62
QY	970	PheMetAsnPheGlnAspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAla	989
DB	63	CTCATGAGCTTTGAGGATGTTCTGTAGTATTTCAGCTGTGTCTTGGTCCCATGGCC	122
QY	990	ValGlyGlnValSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis	1009
DB	123	GTGGGGCAAGTCAGTTCAITTCCTGACTATGCCAAAGCCAAATAATATCAGCAGCCAC	182
QY	1010	ValIleMetIleLeuLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys	1029
DB	183	ATCATCATGATCATTTGAAAGAACCCCTTTGATTCAGCTAGCTAGCAGCAGGAGCCCTAATG	242
QY	1030	ProAsnThrLeuGluGlyAsnValThrPheAsnGluValPheAsnTyrProThrArg	1049
DB	243	CCGAACACATGGAAGAAATGTCACATTTGGTGAAGTTGTATCAACTATCCACCCGA	302
QY	1050	ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAla	1069
DB	303	CCGGACATCCAGTGTTCAGGAGCTGAGCCTGGAGGTGAAGAGGCCAGCGCTGGCT	362
QY	1070	LeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyr	1089
DB	363	CTGGTGGCAGCAGTGGCTGTGGGAAGAGCAGCAGTGGTCCAGCTCTGGAGCGGTCTAC	422
QY	1090	AspProLeuAlaGlySerValLeuIleAspGlyLysGluLeuLysHisLeuAsnValGln	1109
DB	423	GACCCCTGGCAGGAAAGTGTCTGTGATGGCAAGAAATAAAGGAGCTGAATGTTACG	482
QY	1110	TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle	1129
DB	483	TGGCTCCGAGCACACCTGGGCATCGTGTCCAGGAGGCCATCTCTGTGACTCGACATT	542
QY	1130	AlaGluAsnIleAlaTyrGly	1136
DB	543	GCTGAGAACATTCGCTATGGA	563

RESULT 15
BM052878
LOCUS
DEFINITION ie67e03.y3 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
cDNA 5' similar to SW.MDR1_HUMAN P08183 MULTIDRUG RESISTANCE
PROTEIN 1 ; mRNA sequence.

ACCESSION
VERSION BM052878
KEYWORDS EST.
SOURCE
ORGANISM human

REFERENCE
AUTHORS
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Seearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hiller,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.

JOURNAL
COMMENT
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
High quality sequence stop: 430.

FEATURES
source

Location/Qualifiers
1. .563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 136 a 132 c 153 g 142 t
ORIGIN

Alignment Scores:
Pred. No.: Length: 563
Score: 889.00 Matches: 171
Percent Similarity: 96.26% Conservative: 9
Best Local Similarity: 91.44% Mismatches: 7
Query Match: 13.73% Indels: 0
DB: 10 Gaps: 0

US-09-672-725C-27 (1-1281) x BM052878 (1-563)

QY	950	MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu	969
DB	3	ATGATGATATTTTCCCTATGCTGGATGTTTCCGGTTTGGAGCCTACTTGGTGGCACAATAA	62
QY	970	PheMetAsnPheGlnAspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAla	989
DB	63	CTCATGAGCTTTGAGGATGTTCTGTAGTATTTCAGCTGTGTCTTGGTCCCATGGCC	122
QY	990	ValGlyGlnValSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis	1009
DB	123	GTGGGGCAAGTCAGTTCAITTCCTGACTATGCCAAAGCCAAATAATATCAGCAGCCAC	182
QY	1010	ValIleMetIleLeuLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys	1029
DB	183	ATCATCATGATCATTTGAAAGAACCCCTTTGATTCAGCTAGCTAGCAGCAGGAGCCCTAATG	242
QY	1030	ProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArg	1049
DB	243	CCGAACACATGGAAGAAATGTCACATTTGGTGAAGTTGTATCAACTATCCACCCGA	302
QY	1050	ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAla	1069
DB	303	CCGGACATCCAGTGTTCAGGAGCTGAGCCTGGAGGTGAAGAGGCCAGCGCTGGCT	362
QY	1070	LeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyr	1089
DB	363	CTGGTGGCAGCAGTGGCTGTGGGAAGAGCAGCAGTGGTCCAGCTCTCTGGAGCGGTCTAC	422

QY 1090 AspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGln 1109
 Db 423 GACCCCTTGGCAGGGAAGTCTGCTTGATGGCAAGAAATAAGCGACTGAATGTTCA 482
 QY 1110 TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle 1129
 Db 483 TGGCTCCGAGCACACCTGGGCATCGTCCAGGAGCCATCCTGTTGACTGCAGCAT 542
 QY 1130 AlaGluAsnIleAlaTyrGly 1136
 Db 543 GCTGAGAACATTGCCTATGGA 563

Search completed: November 7, 2002, 07:29:43
 Job time : 2253.96 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 17:45:54 ; Search time 31.4697 Seconds
(without alignments)
4517.824 Million cell updates/sec

Title: US-09-672-725C-4

Perfect score: 6467

Sequence: 1 MPEGRKGSAGKNEFKMGK.....LLAQKGIYFSMISVQAGAKR 1280

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*

2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*

3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*

4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*

5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*

6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*

7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*

8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*

9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*

10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*

11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*

12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*

13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*

14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*

15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*

16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*

17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*

18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*

19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*

20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*

21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*

22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6467	100.0	1280	22	Dog P-glycoprotein
2	6467	100.0	1280	22	Dog P-glycoprotein
3	6425.5	99.4	1281	22	Dog P-glycoprotein
4	6425.5	99.4	1281	22	Dog P-glycoprotein
5	6425.5	99.4	1281	22	Dog P-glycoprotein
6	6420.5	99.3	1281	22	Dog P-glycoprotein
7	5859.5	90.6	1280	18	Human multidrug re
8	5859.5	90.6	1280	21	Human wild-type mu
9	5859.5	90.6	1280	22	Human MDR1. Homo
10	5855.5	90.5	1280	22	Cynomolgous monke
11	5853	90.5	1283	22	Cynomolgous monke

12	5850.5	90.5	1280	21	AA58187	Human G185V mutant
13	5850.5	90.5	1280	22	AA81066	Human P-glycoprote
14	5850.5	90.5	1280	22	AAE00306	Human P-glycoprote
15	5847.5	90.4	1280	8	AA70452	Sequence encoded b
16	5844.5	90.4	1280	11	AA804868	Protein encoded by
17	5843	90.4	1279	22	AA81067	Human P-glycoprote
18	5843	90.4	1279	22	AAE00307	Human P-glycoprote
19	5837.5	90.3	1280	15	AA63624	Human P-glycoprote
20	5834.5	90.2	1280	14	AA844297	Human P-glycoprote
21	5834.5	90.2	1280	22	AAU04347	Sequence encoded b
22	5834	90.2	1279	19	AAW80294	Human multidrug re
23	5823.5	90.0	1280	19	AAW48997	Mutated human P-gl
24	5817.5	90.0	1280	19	AAW48998	Wild-type human P
25	5817.5	90.0	1280	19	AAW48999	N-terminal single
26	5811.5	89.9	1280	19	AAW49000	C-terminal single
27	5638.5	87.2	1276	21	AA58189	Human P glycoprote
28	5621.5	86.9	1272	21	AA770597	Marine multidrug r
29	5611.5	86.8	1272	21	AA770596	Rat multidrug resi
30	5304.5	82.0	1276	14	AA835199	Rat multidrug resi
31	5304.5	82.0	1276	14	AA835198	Mouse multidrug re
32	5296	81.9	1275	21	AA78879	Marine multidrug r
33	5296	81.9	1275	22	AA60409	Rat mdrlb2 (multis
34	2797	43.3	1275	20	AAW82594	Rat mdrlb2 multidr
35	2685.5	41.5	1313	22	AB63904	H. contortus PEP-A
36	2554	39.5	1279	22	AB60234	Drosophila melanog
37	2511.5	38.8	1349	20	AA716434	Drosophila melanog
38	2491	38.5	1302	22	AB64919	Multiple drug resi
39	2482	38.4	1334	20	AAW02630	Drosophila melanog
40	2432.5	37.6	1408	19	AAW62871	Protein encoded by
41	2423	37.5	1286	21	AA639101	Multiple drug resi
42	2404.5	37.2	1333	22	AB62495	Arabidopsis thalia
43	2396	37.0	1243	21	AA639102	P. chrysogenum ABC
44	2395	37.0	1320	22	AB64922	Arabidopsis thalia
45	2384	36.9	1254	21	AA92173	Drosophila melanog
						A. thaliana ATPAC,

ALIGNMENTS

RESULT 1

AA881068

ID AA881068 standard; Protein; 1280 AA.

AC AA881068;

DT 25-JUN-2001 (first entry)

XX Dog P-glycoprotein SEQ ID 7.

DE Dog P-glycoprotein

XX Cynomolgous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;

KW efflux pump; dog.

XX Canis familiaris.

OS WO200123565-A1.

PN 05-APR-2001.

PD 28-SEP-2000; 2000WO-US26592.

PF 28-SEP-1999; 99US-0156921.

PR 12-OCT-1999; 99US-0158818.

XX (GENT-) GENTEST CORP.

XX Stocker PJ, Steimel-Crespi DT, Crespi CL;

XX WPI; 2001-316136/33.

XX Novel isolated nucleic acid encoding cynomolgous monkey P-glycoprotein

PT (PGP) and homologous PGP polypeptides are useful for predicting

PT bioavailability of compound and increasing PGP transporter activity in

PT cell

XX
PC
Claim 8: Page 73-76; 84pp; English.

This invention relates to a polynucleotide sequence encoding a cynomolgous monkey p-glycoprotein (PGP), and an allelic variant of the PGP protein. PGP, also known as multidrug transporter, MDR1 is a member of the ABC transporter superfamily. The enzyme serves as an efflux pump exporting small molecules across the cell membrane. The invention includes a cynomolgous monkey (Macaca fascicularis) PGP coding sequence and protein, and also that of an allelic variant. The PGP polynucleotide sequence is useful for increasing PGP transporter activity in a cell. Antisense sequences of the cDNA are useful for inhibiting PGP transport activity in a mammalian cell. They may also be used for increasing the bioavailability of a drug. The present sequence represents p-glycoprotein from a dog.

XX	Sequence	1280 AA;
S0		

Query Match	100.0%;	Score 6467;	DB 22;	Length 1280;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1280:	Conservative	0;	Mismatches	0;
		0;	Indels	0;
			Gaps	0;

Qy	1	MDPGRGKSAEKNFWMKKSKKEKKPKVSTFAMFRYSNWLDRULMVLGVGTMAAIIH	60
Dd	1	MDPEGKSGAEKNFMKMGGKSKKEKKPKVSTFAMFRYSNWLDRULMVLGVGTMAAIIH	60
Qy	61	GAAPLMLLVFGNMDSFANAGISRNKTPPVIIINESITNTQHFINHLEEMTTVAYYIS	120
Dd	61	GAAPLMLLVFGNMDSFANAGISRNKTPPVIIINESITNTQHFINHLEEMTTVAYYIS	120
Qy	121	GIGAGVLVAAYIQVSFWCLAAAGROILLIKRKQFFHAMRQEIGFDVDVGELNTRLTDVV	180
Dd	121	GIGAGVLVAAYIQVSFWCLAARGROILLIKRKQFFHAMRQEIGFDVDVGELNTRLTDVV	180
Qy	181	SKINEGIDKVGMEFQSJATFTTFFGIVGFTPGWKLLVLILAISPVLGLSAAIWAAILSSF	240
Dd	181	SKINEGIDKVGMEFQSJATFTTFFGIVGFTPGWKLLVLILAISPVLGLSAAIWAAILSSF	240
Qy	241	TDKELLAYAKAGAAVEVLAARTVIAFGGOKKELERYNNLEEAKRGIKKAITANISI	300
Dd	241	TDKELLAYAKAGAAVEVLAARTVIAFGGOKKELERYNNLEEAKRGIKKAITANISI	300
Qy	301	GAAPFLLIYASVALAFWGTSLVLSSSEYTIQOVLTVPFSVLIGNAFSIGOASPSTIEAFANAR	360
Dd	301	GAAPFLLIYASVALAFWGTSLVLSSSEYTIQOVLTVPFSVLIGNAFSIGOASPSTIEAFANAR	360
Qy	361	GAAYEIFKIINDKPSIDSYSKSGHKPNIKGNLFKNVHFSPSRKEVKILKGLNLKVQS	420
Dd	361	GAAYEIFKIINDKPSIDSYSKSGHKPNIKGNLFKNVHFSPSRKEVKILKGLNLKVQS	420
Qy	421	GOTVALVNSGCGKSTTVQLMORLYDPTDGMVCIDGODIRTINVRLHREITGVVSGPEVL	480
Dd	421	GOTVALVNSGCGKSTTVQLMORLYDPTDGMVCIDGODIRTINVRLHREITGVVSGPEVL	480
Qy	481	FATTIAENIRYGRENVTWDETEKAVKANAYDFTMKLPNKFDTLVBGERGARLSGGQKRI	540
Dd	481	FATTIAENIRYGRENVTWDETEKAVKANAYDFTMKLPNKFDTLVBGERGARLSGGQKRI	540
Qy	541	ATARALVRNPKTLLIDEATSALDTESEAVVOVALDKARKGRTTIVAHLRSTVRNADVIA	600
Dd	541	ATARALVRNPKTLLIDEATSALDTESEAVVOVALDKARKGRTTIVAHLRSTVRNADVIA	600
Qy	601	GFDGDVIVEKGHNDELMEKGIYPFKLVTMQTRGNEIELENATGESKSESDALEMSPKDGS	660
Dd	601	GFDGDVIVEKGHNDELMEKGIYPFKLVTMQTRGNEIELENATGESKSESDALEMSPKDGS	660
Qy	661	SSLIKRRSTRRSIHAPQGDRKLGTKEDLNENVSVSFWRLLKLNSTEWPFVYVGFICAI	720
Dd	661	SSLIKRRSTRRSIHAPQGDRKLGTKEDLNENVSVSFWRLLKLNSTEWPFVYVGFICAI	720
Qy	721	IAGGLOPAESIIFSRITIGTFTRDEDPETKRONSNMFVLVLVGLIISFIITFFLQGFTEGK	780
Dd	721	INGGLOPAESIIFSRITIGTFTRDEDPETKRONSNMFVLVLVGLIISFIITFFLQGFTEGK	780

xx * The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC "knockout" phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (PGP). The
 CC PGP enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.
 xx
 xx Sequence 1280 AA;
 SQ
 Query Match 100.0%; Score 6467; DB 22; Length 1280;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDPEGRGSAEKNFWMGKSKKKEKPTVSTFAMFRYSNWLDRKLYMLVGTMAAIIH 60
 Db 1 MDPEGRGSAEKNFWMGKSKKKEKPTVSTFAMFRYSNWLDRKLYMLVGTMAAIIH 60
 QY 61 GAALPLMLVFGNNTDSFANAGISRNKTPPVINESITNTQHFHNLBEEMTYAYYS 120
 Db 61 GAALPLMLVFGNNTDSFANAGISRNKTPPVINESITNTQHFHNLBEEMTYAYYS 120
 QY 121 GIGAGVLVAAYIQVSWFCLAAAGQILKIRKOPFHATMRQIEGDFVDHVGELNRLTDDV 180
 Db 121 GIGAGVLVAAYIQVSWFCLAAAGQILKIRKOPFHATMRQIEGDFVDHVGELNRLTDDV 180
 QY 181 SKINEGIDKVGMPFOSIAFTFTGFTVGTGPKWKLVLILAIKSPVLSAAIWKILTSF 240
 Db 181 SKINEGIDKVGMPFOSIAFTFTGFTVGTGPKWKLVLILAIKSPVLSAAIWKILTSF 240
 QY 241 TDKELLAYAKAGAAVEEVLAAITVTAFGGOKKELRYNKNLEAKRIGKKAITANISI 300
 Db 241 TDKELLAYAKAGAAVEEVLAAITVTAFGGOKKELRYNKNLEAKRIGKKAITANISI 300
 QY 301 GAAFLIYASALAFWYGTSLVLSSEYTIQVLTFFSVLIGAFSIGOASPSIEAFANAR 360
 Db 301 GAAFLIYASALAFWYGTSLVLSSEYTIQVLTFFSVLIGAFSIGOASPSIEAFANAR 360
 QY 361 GAAEYFKIIDNKPISIDSYSGHGPNDIKGNLFKNVHFSYPSRKEVKILKGLNKLKVS 420
 Db 361 GAAEYFKIIDNKPISIDSYSGHGPNDIKGNLFKNVHFSYPSRKEVKILKGLNKLKVS 420
 QY 421 GQTVLVGNSGCGKSTVQLMQRLDPTDGMVCIDQDITRTNVRHLREITGVVQEPVL 480
 Db 421 GQTVLVGNSGCGKSTVQLMQRLDPTDGMVCIDQDITRTNVRHLREITGVVQEPVL 480
 QY 481 FATTIAENIRYGRNVMTDEIEKAYKEANAYDFIMKLPNKFTLVGERGSLGGOKORI 540
 Db 481 FATTIAENIRYGRNVMTDEIEKAYKEANAYDFIMKLPNKFTLVGERGSLGGOKORI 540
 QY 541 ATARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIA 600
 Db 541 ATARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIA 600
 QY 601 GFDGIVIVEKGNHDELMKEGILYFLVMTQTRGNEIELENATGESKSDALEMSPKDSG 660
 Db 601 GFDGIVIVEKGNHDELMKEGILYFLVMTQTRGNEIELENATGESKSDALEMSPKDSG 660
 QY 661 SSLIKRRSTRRSIHAPOQDRKLGTKEDLNENVPVSVFWRILKLNSTWPFVVGIFCAI 720
 Db 661 SSLIKRRSTRRSIHAPOQDRKLGTKEDLNENVPVSVFWRILKLNSTWPFVVGIFCAI 720

QY 721 INGGLOPAPSIIFSRIGIIFTRDEDETPETKRONSMFVFLVLGIISFTFFLQGFTEGK 780
 Db 721 INGGLOPAPSIIFSRIGIIFTRDEDETPETKRONSMFVFLVLGIISFTFFLQGFTEGK 780
 QY 781 AGSILTKRLRYMVFRRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKGAGSRLAVITON 840
 Db 781 AGSILTKRLRYMVFRRSMLRQDVSFDDPKNTTGALTTRLANDAAQVKGAGSRLAVITON 840
 QY 841 IANLGTGIIISLYGWQLTLLLAIVPIIAIAGVEMKMLSGOALKDKKELEGAGKIATE 900
 Db 841 IANLGTGIIISLYGWQLTLLLAIVPIIAIAGVEMKMLSGOALKDKKELEGAGKIATE 900
 QY 901 AIENFTVVSLLTREQKFEYMYAQSLQVYPYRNSLRKAHIFGVFSFSTQAMMYSYACGCRF 960
 Db 901 AIENFTVVSLLTREQKFEYMYAQSLQVYPYRNSLRKAHIFGVFSFSTQAMMYSYACGCRF 960
 QY 961 GAYLVANEFMNFQDVLVLSAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMIEKSPILID 1020
 Db 961 GAYLVANEFMNFQDVLVLSAIVFGAMAVGVSSFAPDYAKAKVSAAHVIMIEKSPILID 1020
 QY 1021 SYSPHGLKPNTELEGNVTNEVFNTPRDPVLPVQLGLSLEVKKGOTLALVSSGCGKSTV 1080
 Db 1021 SYSPHGLKPNTELEGNVTNEVFNTPRDPVLPVQLGLSLEVKKGOTLALVSSGCGKSTV 1080
 QY 1081 VOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHLGIVISOEPILEDCSAENIAYGDNRSV 1140
 Db 1081 VOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHLGIVISOEPILEDCSAENIAYGDNRSV 1140
 QY 1141 VSHEEIMQAAKENIHHFETLPEKYNTRVGDKGTLQSGGQKORIAIARALVRPHILL 1200
 Db 1141 VSHEEIMQAAKENIHHFETLPEKYNTRVGDKGTLQSGGQKORIAIARALVRPHILL 1200
 QY 1201 DEATSALDTESEKVVQEAALDKARREGTCIVIAHRLSTONADLIIVFONGVKVKEHGHQ 1260
 Db 1201 DEATSALDTESEKVVQEAALDKARREGTCIVIAHRLSTONADLIIVFONGVKVKEHGHQ 1260
 QY 1261 LLAQKGIYFMSISVQAGAKR 1280
 Db 1261 LLAQKGIYFMSISVQAGAKR 1280
 RESULT 3
 AAEE0308
 ID AAEE0308 standard; Protein; 1281 AA.
 XX
 AC AAEE0308;
 XX
 DT 13-JUN-2001 (first entry)
 XX
 DE Dog P-glycoprotein (PGP) allelic variant (Genotype A).
 XX
 KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
 KW MDR1; drug bioavailability; transgenic animal; genetic model.
 XX
 OS Canis familiaris.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 197 /note= "His of GenotypeC substituted by Gln"
 XX
 PN WO200123540-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26767.
 XX
 PR 28-SEP-1999; 99US-0156510.
 XX
 PA (GENT-) GENTEST CORP.
 XX
 PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 XX
 DR WPI; 2001-235373/24.

DR N-PSDB; AAD03504.

XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -

PS Claim 17; Page 91-93; l1lpp; English.

XX The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDR1) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
CC This sequence is also referred as Genotype A protein. The PGP
CC enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.

XX Sequence 1281 AA;

Query Match 99.4%; Score 6425.5; DB 22; Length 1281;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1273; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 1 MDEGRKGSAGNFKWKKSKK-EKKEKKPTVSTFAMFRYSNWLDRLYMLVGTMAAII 59
DB 1 MDEGRKGSAGNFKWKKSKKKEKKPTVSTFAMFRYSNWLDRLYMLVGTMAAII 60
QY 60 HGAALPLMLVFGNMTDSFANAGISRNKTPPVIINESITNTQHFHNHLEEMTYVYY 119
DB 61 HGAALPLMLVFGNMTDSFANAGISRNKTPPVIINESITNTQHFHNHLEEMTYVYY 120
QY 120 SGIGAGVLVAAYIQVSPFWCLAAQRILKIRKQFFHAIMRQEIQFVDVHVGELNLTDD 179
DB 121 SGIGAGVLVAAYIQVSPFWCLAAQRILKIRKQFFHAIMRQEIQFVDVHVGELNLTDD 180
QY 180 VSKNEGIGDKVGMFFOSIATFTGTPGKWLTVILAIISPVGLSAAIWAAILSS 239
DB 181 VSKNEGIGDKVGMFFOSIATFTGTPGKWLTVILAIISPVGLSAAIWAAILSS 240
QY 240 FTDKELLAYAKAGAVEVLAAIRTVIAFGQKKELERYNKNLEBAKRGIGKKAITANIS 299
DB 241 FTDKELLAYAKAGAVEVLAAIRTVIAFGQKKELERYNKNLEBAKRGIGKKAITANIS 300
QY 300 IGAFLIYIAYALAFWYGTSLVLSSEYTGVLTVFFSVLIGAFSIGQASPSIEAFANA 359
DB 301 IGAFLIYIAYALAFWYGTSLVLSSEYTGVLTVFFSVLIGAFSIGQASPSIEAFANA 360
QY 360 RGAAYEIFKIIDNKPSIDSYSKSGKPDNKNLEFNHVSFSPSRKEVKILKGLNLKVQ 419
DB 361 RGAAYEIFKIIDNKPSIDSYSKSGKPDNKNLEFNHVSFSPSRKEVKILKGLNLKVQ 420
QY 420 SGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDRTINVRHLREITGVVSEPV 479
DB 421 SGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDRTINVRHLREITGVVSEPV 480
QY 480 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDPLVGERGARGSGGQKOR 539
DB 481 LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDPLVGERGARGSGGQKOR 540
QY 540 IAIARALVRNPKILLDEATSEALDTESEAVVQVVALDKARKGRITIVIAHRLSTVRNADVI 599
DB 541 IAIARALVRNPKILLDEATSEALDTESEAVVQVVALDKARKGRITIVIAHRLSTVRNADVI 600

QY 600 AGFDGIVIVEKGNHDELMKEKGIYFKLVMTQTRGNEIELEENATGESKESDALEMSPKDS 659
DB 601 AGFDGIVIVEKGNHDELMKEKGIYFKLVMTQTRGNEIELEENATGESKESDALEMSPKDS 660
QY 660 GSSLIKRRSTRSRSHAPQGDQRKLGTYKEDLNENVPVSFWRIKLKLNSTWPFVVGIFCA 719
DB 661 GSSLIKRRSTRSRSHAPQGDQRKLGTYKEDLNENVPVSFWRIKLKLNSTWPFVVGIFCA 720
QY 720 IINGLOPATSIFSRIGITRDEDEPETKRONSMFVFLVLGIIISFTFFLQGTFFG 779
DB 721 IINGLOPATSIFSRIGITRDEDEPETKRONSMFVFLVLGIIISFTFFLQGTFFG 780
QY 780 KAGEILTKRLRYMVRSMRLRODVSFDDPNNTGALTTRLANDAAQVKGAGISRLAVITQ 839
DB 781 KAGEILTKRLRYMVRSMRLRODVSFDDPNNTGALTTRLANDAAQVKGAGISRLAVITQ 840
QY 840 NIANLGTIIISLIYGWQLTLLAIPIAIAAGVEMKMLSGQALKDKKKELEGAGKIAT 899
DB 841 NIANLGTIIISLIYGWQLTLLAIPIAIAAGVEMKMLSGQALKDKKKELEGAGKIAT 900
QY 900 EAIENFRTVSVLTREQKFEYMAQSLQVVPYRNSLRKAHIFGVSPSITQAMMYFSYACGFR 959
DB 901 EAIENFRTVSVLTREQKFEYMAQSLQVVPYRNSLRKAHIFGVSPSITQAMMYFSYACGFR 960
QY 960 FGAYLVANEFNFODVLLVFSIAIVFGAMAYGOVSSFPADYAKAKVSAAHVIMIEKSPLI 1019
DB 961 FGAYLVANEFNFODVLLVFSIAIVFGAMAYGOVSSFPADYAKAKVSAAHVIMIEKSPLI 1020
QY 1020 DSYSPHGLKPNLTGNTVTFNEWENYPTRPDIPVLQGLSLEVKKGQTLALVGGSGCGKST 1079
DB 1021 DSYSPHGLKPNLTGNTVTFNEWENYPTRPDIPVLQGLSLEVKKGQTLALVGGSGCGKST 1080
QY 1080 VVQLLEFYDPLAGSLVDIGKEIKHLNVQWLRHAHLGIVSQEPILFDCSIAENIAYGNSR 1139
DB 1081 VVQLLEFYDPLAGSLVDIGKEIKHLNVQWLRHAHLGIVSQEPILFDCSIAENIAYGNSR 1140
QY 1140 VVSHEEIMQAAKEANIHFIETLPEKYNTRYGDKGTQLSGGOKORIAIARALVRQPHILL 1199
DB 1141 VVSHEEIMQAAKEANIHFIETLPEKYNTRYGDKGTQLSGGOKORIAIARALVRQPHILL 1200
QY 1200 LDEATSAIDTSEKVVQVQALDKAREGRTCIIVIAHRLSTIQNALDIVFQNGKVEHGHQ 1259
DB 1201 LDEATSAIDTSEKVVQVQALDKAREGRTCIIVIAHRLSTIQNALDIVFQNGKVEHGHQ 1260
QY 1260 QLLAOKGIYFSMISVOAGAKR 1280
DB 1261 QLLAOKGIYFSMISVOAGAKR 1281

RESULT 4

AAE00309 standard; Protein; 1281 AA.

XX AAE00309;

XX 13-JUN-2001 (first entry)

XX Dog P-glycoprotein (PGP) allelic variant (Genotype B).

XX Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;

XX MDR1; drug bioavailability; transgenic animal; genetic model.

XX Canis familiaris.

XX Key Location/Qualifiers

XX Misc-difference 25 /note= "Asn of GenotypeC substituted by Lys"

XX Misc-difference 197 /note= "His of GenotypeC substituted by Gln"

XX WO200123540-A2.

XX 05-APR-2001.

PD

XX PF- 28-SEP-2000; 2000WO-US26767.
 XX PR 28-SEP-1999; 99US-0156510.
 XX FA (GENT-) GENTEST CORP.
 XX PI Stocker PJ, Stelmel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 XX DR WPI: 2001-235373/24.
 XX DR N-PSDB; AAD03505.
 XX PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 XX PT for determining the bioavailability of drugs and for screening for dog
 XX XX PGP inhibitors -
 XX PS Claim 17; Page 99-102; 111pp; English.
 XX CC The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
 CC This sequence is also referred as Genotype B protein. The PGP
 CC enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.
 XX SQ Sequence 1281 AA;
 Query Match 99.4%; Score 6425.5; DB 22; Length 1281;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 1273; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
 QY 1 MDPEGRKSGSAENFWMKGRKS-KKEKKEKPTVSTFAMFYSNWLDRMLVGTMAAII 59
 DB 1 MDPEGRKSGSAENFWMKGRKS-KKEKKEKPTVSTFAMFYSNWLDRMLVGTMAAII 60
 QY 60 HGAALPLMLVFGNMTDSFANAGISRKNKTPPVINISITNNTQHFINHLEEMTTAYYY 119
 DB 61 HGAALPLMLVFGNMTDSFANAGISRKNKTPPVINISITNNTQHFINHLEEMTTAYYY 120
 QY 120 SGIGAGVLVAAYIQVSWFCLAAQRQILKIRKQFFHAIMRQEGWFDVHDVGELNTRLTDD 179
 DB 121 SGIGAGVLVAAYIQVSWFCLAAQRQILKIRKQFFHAIMRQEGWFDVHDVGELNTRLTDD 180
 QY 180 VSKINEGDKGVMTFQSIATFTFTGFTVGTGKTLVLILAIISPVILGSAALWAKILSS 239
 DB 181 VSKINEGDKGVMTFQSIATFTFTGFTVGTGKTLVLILAIISPVILGSAALWAKILSS 240
 QY 240 FTDKELLAYAKAGAAVEVLAARTVIAFGGQKKELEKERNKNLEAKRIGIKKAITANIS 299
 DB 241 FTDKELLAYAKAGAAVEVLAARTVIAFGGQKKELEKERNKNLEAKRIGIKKAITANIS 300
 QY 300 IGAALLIYASALAFWYGTSLVLSSEYITIGVLTFFSVLIGAFSIGOASPSIEAFANA 359
 DB 301 IGAALLIYASALAFWYGTSLVLSSEYISIGVLTFFSVLIGAFSIGOASPSIEAFANA 360
 QY 360 RGAAYEIFKIIDNKPSIDYSKSGHRPDNIKGNLEPKNFHFSYPSRKEVKILKGLNKVQ 419
 DB 361 RGAAYEIFKIIDNKPSIDYSKSGHRPDNIKGNLEPKNFHFSYPSRKEVKILKGLNKVQ 420
 QY 420 SGQTVALVGNCGCKSTTVOLMQRLYDPTDGMVCIDQDQIRTNVRLHREITGVVSQEPV 479

DB 421 SGQTVALVGNCGCKSTTVOLMQRLYDPTDGMVCIDQDQIRTNVRLHREITGVVSQEPV 480
 QY 480 LFATTIAENIRYGRNVTMDIEIEKAYKEANAYDFIMKLPNKFTLVGERGARGLSGQKQR 539
 DB 481 LFATTIAENIRYGRNVTMDIEIEKAYKEANAYDFIMKLPNKFTLVGERGARGLSGQKQR 540
 QY 540 IATARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 599
 DB 541 IATARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
 QY 600 AGFDDGVIVEKGNHDELMKEKGIYFKLVMTOTRGNEIELENATGESKSDALEMPKDS 659
 DB 601 AGFDDGVIVEKGNHDELMKEKGIYFKLVMTOTRGNEIELENATGESKSDALEMPKDS 660
 QY 660 GSSLIKRRSTRRSIHAPQGOQDRKLTGEDLNENVPVSFWRLIKLNSTWPFVVGIFCA 719
 DB 661 GSSLIKRRSTRRSIHAPQGOQDRKLTGEDLNENVPVSFWRLIKLNSTWPFVVGIFCA 720
 QY 720 IINGGLQPAFSIIISRIIGITFDEDEPETKRONSMFSLVLVLGLIISITFFLQGFTEG 779
 DB 721 IINGGLQPAFSIIISRIIGITFDEDEPETKRONSMFSLVLVLGLIISITFFLQGFTEG 780
 QY 780 KAGEIITKRLRYMFRSMLRQDVSWFDDPKNTGALTTRLANDAAQVKAIGSLAVITQ 839
 DB 781 KAGEIITKRLRYMFRSMLRQDVSWFDDPKNTGALTTRLANDAAQVKAIGSLAVITQ 840
 QY 840 NIANLGTGIIISLIYQWQJTLTLLAIVPIIAIAGVVMKMLSGQALKDKELEGAGKIAT 899
 DB 841 NIANLGTGIIISLIYQWQJTLTLLAIVPIIAIAGVVMKMLSGQALKDKELEGAGKIAT 900
 QY 900 EATENERTVVSUTREOKFEYMYAQSLQVYRNSLRKAHIFGVFSITQAMMTFSYAGCFR 959
 DB 901 EATENERTVVSUTREOKFEYMYAQSLQVYRNSLRKAHIFGVFSITQAMMTFSYAGCFR 960
 QY 960 FGAYLVANEFMNFQDVLVFSALVFGAMAVGVSSFPADYAKAKYSAAHVIMIEKSPLI 1019
 DB 961 FGAYLVANEFMNFQDVLVFSALVFGAMAVGVSSFPADYAKAKYSAAHVIMIEKSPLI 1020
 QY 1020 DSYSPHGLKPNLEGNVTNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSSGCGKST 1079
 DB 1021 DSYSPHGLKPNLEGNVTNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSSGCGKST 1080
 QY 1080 VVQLLERFYDPLAGSVLIDGKEIKHLNVQWLAHILGIVSOEPILEDSCSAENIATYDNR 1139
 DB 1081 VVQLLERFYDPLAGSVLIDGKEIKHLNVQWLAHILGIVSOEPILEDSCSAENIATYDNR 1140
 QY 1140 VVSHEEIMQAKEANIHHFIETLPEKYNTRVGDGKTQSLGGQKQRIATARALVRPHILL 1199
 DB 1141 VVSHEEIMQAKEANIHHFIETLPEKYNTRVGDGKTQSLGGQKQRIATARALVRPHILL 1200
 QY 1200 LDEATSDALDTESEKVVQZALDKAREGRTCTIVIAHRLSTIONADLLIVFQNGKVEHGTQ 1259
 DB 1201 LDEATSDALDTESEKVVQZALDKAREGRTCTIVIAHRLSTIONADLLIVFQNGKVEHGTQ 1260
 QY 1260 QLLAQKGIYFMSIVQAGAKR 1280
 DB 1261 QLLAQKGIYFMSIVQAGAKR 1281
 RESULT 5
 AAE00310
 ID AAE00310 standard; Protein; 1281 AA.
 XX AC AAE00310;
 XX DT 13-JUN-2001 (first entry)
 XX XX Dog P-glycoprotein (PGP) allelic variant (Genotype D).
 XX XX Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
 KW MDR1; drug bioavailability; transgenic animal; genetic model.
 XX OS Canis familiaris.

XX Key Location/Qualifiers
FH Misc-difference 25 /note= "Asn of GenotypeC substituted by Lys"
FT Misc-difference 197 /note= "His of GenotypeC substituted by Gln"
FT Misc-difference 329 /note= "Ser of GenotypeC substituted by Thr"
FT Misc-difference 1148 /note= "Met of GenotypeC substituted by Val"
XX
PN WO200123540-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26767.
XX
XX 28-SEP-1999; 99US-0156510.
XX (GENT-) GENTEST CORP.
PA Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
PI N-PSDB; AAD03506.
XX
XX WPI; 2001-235373/24.
DR N-PSDB; AAD03506.
XX
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
XX Claim 17; Page 108-110; 111pp; English.
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDR1) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
CC This sequence is also referred as Genotype D protein. The PGP
CC enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.
XX
SQ Sequence 1281 AA;
Query Match 99.4%; Score 6425.5; DB 22; Length 1281;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1273; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
QY 1 MDPEGGRGSAENFWKMGKKS-KKEKKEKPTVSTFAMPRYSNWLDRYMLVGTMAAII 59
DB 1 MDPEGGRGSAENFWKMGKKS-KKEKKEKPTVSTFAMPRYSNWLDRYMLVGTMAAII 60
QY 60 HGAALPLMLVFGNMTDSFANAGISRKNKTPPVINISITNNTQHFNLHLEEMTTTAYYY 119
DB 61 HGAALPLMLVFGNMTDSFANAGISRKNKTPPVINISITNNTQHFNLHLEEMTTTAYYY 120
QY 120 SGIGAGVLVAAYIQVSWFCWLAAGQILKIRKQFFHAIMRQEIOWFDVHVGELNTRLTDD 179
DB 121 SGIGAGVLVAAYIQVSWFCWLAAGQILKIRKQFFHAIMRQEIOWFDVHVGELNTRLTDD 180
QY 180 VSKINEGIGDKVGMFFQSIATFTFTGTVGTPGKLTFLVLAISPVGLSAAIWAAILSS 239
DB 181 VSKINEGIGDKVGMFFQSIATFTFTGTVGTPGKLTFLVLAISPVGLSAAIWAAILSS 240

QY 240 FTDKELLAYAKAGAVAEVLAARTVIAFGGOKKELERYKNKLEAKRIGIKKAITANIS 299
DB 241 FTDKELLAYAKAGAVAEVLAARTVIAFGGOKKELERYKNKLEAKRIGIKKAITANIS 300
QY 300 IGA AFLLIYASYALAFWYGTSLVLSSEYTTIGQVLTVPFVSVLIGAFSIGQASPSIEAFANA 359
DB 301 IGA AFLLIYASYALAFWYGTSLVLSSEYTTIGQVLTVPFVSVLIGAFSIGQASPSIEAFANA 360
QY 360 RGAAYEIFKIIDNKPSIDSYSGHKKPDNIKGNLEFNHVSFSPSRKEVKILKGLNLKVQ 419
DB 361 RGAAYEIFKIIDNKPSIDSYSGHKKPDNIKGNLEFNHVSFSPSRKEVKILKGLNLKVQ 420
QY 420 SGQTVALVGNSSCGCKSTTVQLMQRLYDPTDGMVCIDQDITINVRHLREITVWVSEPV 479
DB 421 SGQTVALVGNSSCGCKSTTVQLMQRLYDPTDGMVCIDQDITINVRHLREITVWVSEPV 480
QY 480 LFATTIAENIRYGRNVTMDIEIAKVEANAYDFIMKLPNKFDTLVGERGARSQKQOR 539
DB 481 LFATTIAENIRYGRNVTMDIEIAKVEANAYDFIMKLPNKFDTLVGERGARSQKQOR 540
QY 540 IAIARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 599
DB 541 IAIARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 600
QY 600 AGFDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKDS 659
DB 601 AGFDDGVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGESKESDALEMSPKDS 660
QY 660 GSSILKRRSTRSIIHAPQGDRLKGTEDLNENVPVSFWIRILKLNSTEMPYFVVGIFCA 719
DB 661 GSSILKRRSTRSIIHAPQGDRLKGTEDLNENVPVSFWIRILKLNSTEMPYFVVGIFCA 720
QY 720 IINGLOPAFSIIIFRSIIGITRDEDETPKQNSNMFVLFVLGLIISFTIFFLOGTFG 779
DB 721 IINGLOPAFSIIIFRSIIGITRDEDETPKQNSNMFVLFVLGLIISFTIFFLOGTFG 780
QY 780 KAGEITLRLRYMWFERSMLRODVSWDFDPKNTGALTTRLANDAAQVKAIGSLAVITQ 839
DB 781 KAGEITLRLRYMWFERSMLRODVSWDFDPKNTGALTTRLANDAAQVKAIGSLAVITQ 840
QY 840 NIANLGTGIIISLIYGWOLTLLLAIVPIIAIAGVVENKMLSGQALKDKKELEGAGKIAT 899
DB 841 NIANLGTGIIISLIYGWOLTLLLAIVPIIAIAGVVENKMLSGQALKDKKELEGAGKIAT 900
QY 900 EAIENFRVVSITRQKFEYMAQSLOVPYRNSLRKAHIFGVFSFISITQAMMYFSYAGCFR 959
DB 901 EAIENFRVVSITRQKFEYMAQSLOVPYRNSLRKAHIFGVFSFISITQAMMYFSYAGCFR 960
QY 960 FGAYLVANEFMNFQDVLVFSAIYFGAMAVGVSVSFADYAKAVSAAHVIMIEKSPLI 1019
DB 961 FGAYLVANEFMNFQDVLVFSAIYFGAMAVGVSVSFADYAKAVSAAHVIMIEKSPLI 1020
QY 1020 DSYSPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVGSSECKST 1079
DB 1021 DSYSPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVGSSECKST 1080
QY 1080 VVQLLERYFDPLAGSVLIDGKEIKHLNVOWLRHILGIVSQBPILFDCSIAENIAYGDNRS 1139
DB 1081 VVQLLERYFDPLAGSVLIDGKEIKHLNVOWLRHILGIVSQBPILFDCSIAENIAYGDNRS 1140
QY 1140 VVSHEEIMOAAKEANHHFETLPEKYNTRVGDGKTQLSGGQKQRIATARALVRQPHILL 1199
DB 1141 VVSHEEIMOAAKEANHHFETLPEKYNTRVGDGKTQLSGGQKQRIATARALVRQPHILL 1200
QY 1200 LDEATSAIDTESEKVVQBALDKAREGRTCIIVIAHRLSTIQNADLIIVFQNGKVEHGTHQ 1259
DB 1201 LDEATSAIDTESEKVVQBALDKAREGRTCIIVIAHRLSTIQNADLIIVFQNGKVEHGTHQ 1260
QY 1260 QLLAOKGIYFSMISVOAGAKR 1280
DB 1261 QLLAOKGIYFSMISVOAGAKR 1281

RESULT 6	
AAE00303	
ID	AAE00303 standard; Protein; 1281 AA.
XX AC	AAE00303;
XX DT	13-JUN-2001 (first entry)
DE DE	Dog (PGP) P-glycoprotein (genotype C) #1.
XX KW	Dog; P-glycoprotein; PGP; multidrug transporter; MDRI;
XX KW	drug bioavailability; transgenic animal; genetic model.
XX OS	Canis familiaris.
XX PN	WO200123540-A2.
XX PD	05-APR-2001.
XX PF	28-SEP-2000; 2000WO-US26767.
XX PR	28-SEP-1999; 99US-0156510.
XX PA	(GENT-) GENTEST CORP.
XX PI	Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX DR	WPI; 2001-235373/24.
XX DR	N-PSDB; RAD03488.
XX PT	New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX PT	for determining the bioavailability of drugs and for screening for dog
XX PS	PGP inhibitors -
XX PS	Claim 17; Page 64-66; 111pp; English.
CC CC	The invention relates to dog P-glycoprotein (PGP) also referred
CC CC	as multidrug transporter (MDRI) and nucleic acids encoding them.
CC CC	The invention also includes fragments and biologically functional
CC CC	variants of dog P-glycoprotein. PGP and their nucleic acids are
CC CC	useful for determining the bioavailability of drugs and for
CC CC	screening PGP inhibitors. They are useful for the diagnosis and
CC CC	treatment of conditions characterised by PGP activity, by
CC CC	reducing or increasing PGP activity in a cell. PGP nucleic acids
CC CC	are used as oligonucleotide probes. Complements of PGP nucleic
CC CC	'acids are useful as antisense oligonucleotides, to induce a PGP
CC CC	'knockout' phenotype. They are used to prepare a non-human
CC CC	transgenic animal, which are valuable as genetic models for
CC CC	human diseases.
CC CC	The present sequence is dog P-glycoprotein (PGP) also referred
CC CC	as genotype C protein. The PGP enzyme functions as an efflux
CC CC	pump exporting small molecules across the cell membrane. This
CC CC	enzyme is a member of the ABC transporter family.
XX SQ	Sequence 1281 AA;
	Query Match 99.3%; Score 6420.5; DB 22; Length 1281;
	Best Local Similarity 99.3%; Pred. No. 0;
	Matches 1272; Conservative 4; Mismatches 4; Indels 1; Gaps
QY 1	MDPEGKKGSAENFWKMGGKSKK-EKKKKKPTVSTFAMFRYSNWLDRILMLVGTMAAII 59
Db 1	MDPEGKKGSAENFWKMGGKSKKKEKKKPTVSTFAMFRYSNWLDRILMLVGTMAAII 60
QY 60	HGAALPLMLVFCNMWTDSPANAGISRNKTFPVIIINESITNTQHFTHHLEEMTYAYYY 119
Db 61	HGAALPLMLVFCNMWTDSPANAGISRNKTFPVIIINESITNTQHFTHHLEEMTYAYYY 119
QY 120	SGICAGVLVAAYIQVSFWCLAGROQLTKRKOFFFAHMQEIGWDVDHVGELNRLTDD 179
Db 121	SGICAGVLVAAYIQVSFWCLAGROQLTKRKOFFFAHMQEIGWDVDHVGELNRLTDD 180
QY 180	VSKINGIGDGVMFPQSIAATFTTGVTGPKWLTLIVLAISVPGLSAAIWAAILSS 239

Db 1261 QLLAQGIYFSMVQAGAKR 1281

RESULT 7
AAW44073
ID AAW44073 standard; protein; 1280 AA.
XX
AC AAW44073;
XX
DT 26-JUN-1998 (first entry)
XX
DE Human multidrug resistance P-glycoprotein MDRL.
XX
KW Human; multidrug resistance P-glycoprotein; MDRL; prokaryotic homologue.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 1..640
FT /note= "MDRL-N from Fig 1"
FT Region 641..1280
FT /note= "MDRL-C from Fig 1"
XX
XX WO9740160-A1.
XX
XX 30-OCT-1997.
XX
XX 24-APR-1997; 97WO-N000216.
XX
XX 24-APR-1996; 96EP-0201094.
XX
XX (UYGR-) RIJUNIV GRONINGEN.
XX
XX Bolhuis H, Konings MN, Van Veen HW, Venema K;
XX WPI; 1997-535844/49.
XX
XX Prokaryotic homologue of human multiple drug resistance protein -
XX used to screen for compounds that inhibit, or avoid, drug resistance
XX
XX Claim 10; Fig 1; 35pp; English.
XX
XX The present invention describes a recombinant or isolated nucleic acid
XX (I), derived from a prokaryotic gene, which encodes at least a specific
XX and/or functional part of a transporter protein (TP), or its
XX derivatives, which has functional and/or structural similarity with the
XX P-glycoprotein (PG) encoded by the human multidrug resistance
XX (MDR)1 gene. The present sequence represents the human MDRL protein,
XX derived from MDRL-N and MDRL-C as shown in the specification in
XX figure 1. (I) is used to express recombinant proteins; its fragments
XX are also useful as probes and primers for detection and amplification
XX of related DNA. The protein produced, or cells expressing them, are
XX used to determine if substances can inhibit, or avoid, MDR proteins,
XX and in a screening method for identifying compounds that inhibit
XX transport of cytotoxic substances from cells. Also, cells with a
XX transmembrane protein, especially where expressed from (I), can provide
XX (additional) MDR, particularly for use as a model system to study
XX mechanisms of action of PG.
XX
SQ Sequence 1280 AA;
Query Match 90.6%; Score 5859.5; DB 18; Length 1280;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 1159; Conservative 59; Mismatches 60; Indels 3; Gaps 3;
Qy 1 MDPEGRKGS-A-EKNFWKMGKSKKKEKPTVSTFAMPRYSNWLDRLYMLVGTMAII 59
Db 1 MDLEDRNGGAKKNFFNKLNNSEKDKKPTVSVFMSFYSNWLDRLYMLVGTMAII 60
Qy 60 HCAALPLAMLVGNMTDSFANAGISRNKTFPVIINESITNTQHFINHLEEMTYAYY 119
Db 61 HGAGLPLAMLVGEMTDIFANAG-NLELMSNITNRSINDTGFNM-LEEDMTRYAYY 118

Qy 120 SGIGAGVLVAAYIQVSWFCWCLAAAGRIKIRKQFFHAIMRQEIFWDFVDHVGELNTRLTDD 179
Db 119 SGIGAGVLVAAYIQVSWFCWCLAAAGRIKIRKQFFHAIMRQEIFWDFVDHVGELNTRLTDD 178
Qy 180 VSKINEGIDGKVMFFOSIATFTFTGFIYVGTGPKWKLTLVLAISPVGLSAAWAKILSS 239
Db 179 VSKINEGIDGKVMFFOSIATFTFTGFIYVGTGPKWKLTLVLAISPVGLSAAWAKILSS 238
Qy 240 FTDKELLAYAKAGAVAEVLAAIIRTAFAGGQKKELERYNKNLEAKRIGIKKAITANIS 299
Db 239 FTDKELLAYAKAGAVAEVLAAIIRTAFAGGQKKELERYNKNLEAKRIGIKKAITANIS 298
Qy 300 IGAAPLLIYASALAFWYGTSLVLSSEYTTGQVLTFFSVLIGAFSIGQASPSIEAFANA 359
Db 299 IGAAPLLIYASALAFWYGTSLVLSSEYTTGQVLTFFSVLIGAFSIGQASPSIEAFANA 358
Qy 360 RGAAYEIFKIIDNKPSIDSYSGHGDNTKGNLEFNHVSFSPSRKEVKILGNLKVQ 419
Db 359 RGAAYEIFKIIDNKPSIDSYSGHGDNTKGNLEFNHVSFSPSRKEVKILGNLKVQ 418
Qy 420 SGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGDINTINVRLREITGVVSOEPV 479
Db 419 SGQTVLVGNSGCGKSTTVQLMQRLYDPTDGMVCDIGDINTINVRLREITGVVSOEPV 478
Qy 480 LFATTIAENIRYGRNVTMDIEKAVKANAYDEIMKLPNKFDTLVGERGARSQKOR 539
Db 479 LFATTIAENIRYGRNVTMDIEKAVKANAYDEIMKLPNKFDTLVGERGARSQKOR 538
Qy 540 IAIARALVRNPKILLDEATSEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 599
Db 539 IAIARALVRNPKILLDEATSEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 598
Qy 600 AGFDDGVIVEKGNHDELMEKGIYFKVTMTQTRNEIELEENATGESKESDALEMSPKDS 659
Db 599 AGFDDGVIVEKGNHDELMEKGIYFKVTMTQTRNEIELEENATGESKESDALEMSPKDS 658
Qy 660 GSSILKRSTRSHAPQODKLGKTKEDLNENVPVSFWRILKLNSEWFFVVGIFCA 719
Db 659 RSSLIRKSTRSRGSAQDKKJSTKEALDESIPVPSFWRILKLNSEWFFVVGIFCA 718
Qy 720 IINGGLQPAFISIFRIIGITRDEDPETKQNSNMFLVLVLGIIISFIFTFQGGTFFG 779
Db 719 IINGGLQPAFISIFRIIGITRDEDPETKQNSNMFLVLVLGIIISFIFTFQGGTFFG 778
Qy 780 KAGEILTTLRLRYMVRPSMLRQVSWFDDPKNTGTALTRLANDAAQVKGATGSLAVITQ 839
Db 779 KAGEILTTLRLRYMVRPSMLRQVSWFDDPKNTGTALTRLANDAAQVKGATGSLAVITQ 838
Qy 840 NTANLGTGIIISLIYGWQLTLLLAIVPIIATAGVEMKMLSGQALKDKKELEGAKIAT 899
Db 839 NTANLGTGIIISLIYGWQLTLLLAIVPIIATAGVEMKMLSGQALKDKKELEGAKIAT 898
Qy 900 EAIENFRTVSVLTREKQFEYMYAQSLQVYRNSLRKAHIFGVSFISITQAMMYFVAGCFR 959
Db 899 EAIENFRTVSVLTREKQFEYMYAQSLQVYRNSLRKAHIFGVSFISITQAMMYFVAGCFR 958
Qy 960 FGAYLVANEFMFQDVLVFSIAIVFGAMAVGVSSFPADYAKAKYAAHVIMIEKSPLI 1019
Db 959 FGAYLVANEFMFQDVLVFSIAIVFGAMAVGVSSFPADYAKAKYAAHVIMIEKSPLI 1018
Qy 1020 DSYSPHGLKPNLTLEGNTVENEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSSGCGKST 1079
Db 1019 DSYSPHGLKPNLTLEGNTVENEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSSGCGKST 1078
Qy 1080 VVOLLERFYDPLAGSVLIDGKEIKHLNVOWLAHGLIVSQEPILFDCSIAENIAGDNSR 1139
Db 1079 VVOLLERFYDPLAGSVLIDGKEIKHLNVOWLAHGLIVSQEPILFDCSIAENIAGDNSR 1138
Qy 1140 VYSHEETMAAKKANTHHTETLPEKYNTRVGDGKQTLSSGQKQRIATARALVRPHILL 1199
Db 1139 VYSHEETMAAKKANTHHTETLPEKYNTRVGDGKQTLSSGQKQRIATARALVRPHILL 1198
Qy 1200 LDEATSEKVSQV 1259

Db - 1199 LDEATSDLTSEKVVQALDKAREGRTCIIVIAHRLSTIQNADLIIVFQNGRVKEGTHQ 1258
QY 1260 QLLAQKGIYFSMISVQAGAKR 1280
Db 1259 QLLAQKGIYFSMISVQAGTKR 1279

RESULT 8
AAV58186
ID AAV58186 standard; Protein; 1280 AA.
XX
AC AAV58186;
XX
DT 14-MAR-2000 (first entry)
DE Human wild-type multidrug resistance-1 (MDR-1) protein.
XX
KW Multidrug resistance; MDR-1; P-glycoprotein;
KW transmembrane efflux pump; haematopoietic stem cell; transduction;
KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW gene therapy; gene replacement; genetic defect; thalassaemia;
KW Gaucher's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW cytokine; wild-type.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 185
FT /note= "This residue is Val in a mutant MDR-1
FT (AAV58187)"
XX
PN W09961589-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11825.
XX
PR 28-MAY-1998; 98US-0086988.
XX
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Sorrentino B, Bunting K;
XX
XX WPI; 2000-072615/06.
DR N-PSDB; AA249332.
XX
PT Ex vivo expansion of hematopoietic stem cells transduced with a
PT sequence encoding human multidrug resistance-1, used for bone marrow
PT transplantation -
XX
PS Claim 10; Page 71-79; 113pp; English.
XX
CC This sequence represents human wild-type multidrug
CC resistance protein MDR-1. MDR-1 is a transmembrane efflux
CC pump, responsible for the export of drugs from certain
CC cells, particularly cancer cells. Wild-type MDR-1 shows
CC increased resistance to etoposide and decreased resistance to vinca
CC alkaloids compared with a mutant form (AAV58187) where the Gly at
CC position 185 is replaced by Val. The invention relates to transducing
CC haematopoietic stem cells with nucleic acid encoding an MDR protein
CC and culturing the modified cells. The modified haematopoietic stem
CC cells are useful in bone marrow transplantation (to reconstitute
CC haematopoietic systems in patients who have undergone chemotherapy or
CC radiation therapy) and in ex vivo gene therapy of genetic defects in
CC cells derived from haematopoietic stem cells, e.g., thalassaemia,
CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
CC cells can also be used to identify factors involved in regulating
CC proliferation and differentiation in haematopoietic stem cells.
CC Haematopoietic stem cells that express MDR-1 will be protected against
CC chemotherapeutic agents, so can be engrafted while the patient is
CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC provides sufficient cells to permit standard biochemical analysis.

CC Overexpression of MDR-1 allows cytokine-driven expansion of
CC haematopoietic stem cells by at least 10-fold compared with a maximum
CC of 4-fold in known procedures.
XX
SQ Sequence 1280 AA;
Query Match 90.6%; Score 5859.5; DB 21; Length 1280;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 1159; Conservative 59; Mismatches 60; Indels 3; Gaps 3;
QY 1 MDPGGGRKGS-A-EKNEFKMGKSKKKEKKKPTVTSFAMFRYSNWLDRILVGVGMAAII 59
Db 1 MDLEGRNGGAKKNPFKLNKSEKDKKPKTVSVFSNFRYSNWLDRILVGVGMAAII 60
QY 60 HGAALPLMMLVFGNMTDFANAGISRNKTPFVPIINESITNNTQHFHINLHEEMTTAYYY 119
Db 61 HGAGLPLMMLVFGNMTDFANAG-NLEDLMSNITNRSINDTGTFFMN-LEEDMTYAYYY 118
QY 120 SGIGAGVLVAAYIOVSFWCLAGROILKIRKOFFHAIMRQEIFGDFVDHVGELNRLTDD 179
Db 119 SGIGAGVLVAAYIOVSFWCLAGROILKIRKOFFHAIMRQEIFGDFVDHVGELNRLTDD 178
QY 180 VSKINEGIDKVGCMFFQSIATFTFTGFTGKWLTLVLAISPVGLSAAIAWAKILSS 239
Db 179 VSKINEGIDKVGCMFFQSIATFTFTGFTGKWLTLVLAISPVGLSAAIAWAKILSS 238
QY 240 FTDKELLAYAKAGAAEVLAAIRTVIAFGGQKKELERYNKNLEAKRIGIKKAITANIS 299
Db 239 FTDKELLAYAKAGAAEVLAAIRTVIAFGGQKKELERYNKNLEAKRIGIKKAITANIS 298
QY 300 IGAAPLLIYASVALAFWYGTSLVLSSEYTGIVLTVFFSVLIGAFSPSPSEAFANA 359
Db 299 IGAAPLLIYASVALAFWYGTSLVLSSEYTGIVLTVFFSVLIGAFSPSPSEAFANA 358
QY 360 RGAAYEIFKIIDNKPSIDSYSGSHKPDNIKGNLFKFNHFSYPSRKEVKILGLNLKVQ 419
Db 359 RGAAYEIFKIIDNKPSIDSYSGSHKPDNIKGNLFKFNHFSYPSRKEVKILGLNLKVQ 418
QY 420 SGQTVLVGNSCGKSTTVQLMQRLYDPTDGMVCIQDQIRTIINVRHLREITGVVSPV 479
Db 419 SGQTVLVGNSCGKSTTVQLMQRLYDPTDGMVCIQDQIRTIINVRHLREITGVVSPV 478
QY 480 LFATTIAENIRYGRNVTMDIEKAVKANAVDFTMKLPNKFDTLVGERGALSGGQKOR 539
Db 479 LFATTIAENIRYGRNVTMDIEKAVKANAVDFTMKLPNKFDTLVGERGALSGGQKOR 538
QY 540 TAIARALVRNPKILLDLDEATSDLTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 599
Db 539 TAIARALVRNPKILLDLDEATSDLTSEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 598
QY 600 AGFDDGVIVEKGNHDELKKEGIYFKLVMTQTRGNEIELENATGESKSDALEMSPKDS 659
Db 599 AGFDDGVIVEKGNHDELKKEGIYFKLVMTQTRGNEIELENATGESKSDALEMSPKDS 658
QY 660 GSSLIKRSTRSRSHAPQODRKLGTKEDLNENVPVSFWRIKLKLNSTENPWFVVGIFCA 719
Db 659 RSSLIKRSTRSRVSGQAQDKLSTKEALDESIPPSFWRIKLKLNTEPWFVVGIFCA 718
QY 720 IINGLOPAFSLIISRIIGITRDEDPETKQNSNMFVFLVLGIISFTIFFLQGFTEG 779
Db 719 IINGLOPAFSLIISRIIGITRDEDPETKQNSNMFVFLVLGIISFTIFFLQGFTEG 778
QY 780 KAGELTKRLRYMFRSMRLQDVSFDDPKNTTCALTTRLANDAAQVKGAGISRLAVITQ 839
Db 779 KAGELTKRLRYMFRSMRLQDVSFDDPKNTTCALTTRLANDAAQVKGAGISRLAVITQ 838
QY 840 NIANLGTGIIISLYIGWQLTLLIAIPIIAIAGVEMKMLSGOALKDKKELSGAGKIAT 899
Db 839 NIANLGTGIIISLYIGWQLTLLIAIPIIAIAGVEMKMLSGOALKDKKELSGAGKIAT 898
QY 900 EAIENFTVSLTREQFEYMYAQSLQVYRNSLRKAHIFGVFSFISITQAMMYFSYAGCFR 959
Db 899 EAIENFTVSLTREQFEYMYAQSLQVYRNSLRKAHIFGVFSFISITQAMMYFSYAGCFR 958

QY 960 FGAYLVANEFMFQDVLVFSVIFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPLI 1019
DB 959 FGAYLVAHKLMFQDVLVFSVIFGAMAVGVSSFPADYAKAKISAHHIIMIEKTPLI 1018
QY 1020 DSYSPHGLKPNTELEGNTVFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKST 1079
DB 1019 DSYSTEGMLPNTELEGNTVFGEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKST 1078
QY 1080 VVQLLEFYDPLAGSVLIDGKEIKHLNVQWLRAHGLGVSQEPILPDCSIAENIAYGDNRS 1139
DB 1079 VVQLLEFYDPLAGSVLIDGKEIKHLNVQWLRAHGLGVSQEPILPDCSIAENIAYGDNRS 1138
QY 1140 VVSHEIQAAKEANHHFIEITLPEKYNTRVGDKTQLSGGQKQRIATARALVROPHILL 1199
DB 1139 VVSQEEIVRAAKEANHHFIESLPNKYSTKVGDKTQLSGGQKQRIATARALVROPHILL 1198
QY 1200 LDEATSALDTESEKVVQVEALDKAREGTCIVIAHRLSTTIONADLIWVFQNGVKKEHGTQ 1259
DB 1199 LDEATSALDTESEKVVQVEALDKAREGTCIVIAHRLSTTIONADLIWVFQNGVKKEHGTQ 1258
QY 1260 QLLAQKGIYFMSVQAGAKR 1280
DB 1259 QLLAQKGIYFMSVQAGTKR 1279

RESULT 9
AAB81959 standard; protein; 1280 AA.
XX AAB81959;
XX
XX
DT 02-JUL-2001 (first entry)
XX
DE Human MDRL.
XX
XX Human; MDRL; multi-drug resistance; cholesterol uptake;
KW hypercholesterolaemia; hypocholesterolaemia; atherosclerosis;
KW coronary artery disease; cerebral vascular disease.
XX
XX Homo sapiens.
OS
PN WO200121762-A2.
XX
XX 29-MAR-2001.
XX
XX 22-SEP-2000; 2000WO-US26099.
XX
XX 23-SEP-1999; 99US-0155819.
XX
XX (BARN-) BARNES-JEWISH HOSPITAL.
PA
PI Stenson WF, Tessner T;
XX
XX WPI; 2001-328100/34.
XX
XX Modulating cellular cholesterol uptake for treating hyper or
PT hypocholesterolemia, involves administering an agent that inhibits or
PT increases the expression of multiple drug resistance-1 -
XX
XX Example 1; Page 43-47; 47pp; English.
PS
XX
XX The present invention describes a method of modulating cholesterol uptake
CC in cells, involving administering an agent capable of inhibiting the
CC multi-drug resistance protein MDRL. This is useful in the prevention and
CC treatment of cholesterol-related diseases, including
CC hypercholesterolaemia, hypocholesterolaemia, atherosclerosis, coronary
CC artery disease and cerebral vascular diseases. The present sequence is
CC the human MDRL protein.
XX
XX Sequence 1280 AA;
SQ
.. Query Match 90.68; Score 5859.5; DB 22; Length 1280;

Best Local Similarity 90.5%; Pred. No. 0; Mismatches 60; Indels 3; Gaps 3;
Matches 1159; Conservative 59;

QY 1 MDPEGGKSGSA-BKNFWMGKCKKKEKKKPTVSVFAMFYSNWLDRLYMLVGTMAAII 59
DB 1 MDLEGRNGGAKKNFKLNKSEKKKPKIVSVFSMFYSNWLDRLYMLVGTMAAII 60
QY 60 HGAALPLMLVFGNMTDSFANAGISRNKTPFVFINESITNTQHFNLHLEEMTTAYYY 119
DB 61 HGAGLPLMLVFGNMTDIFANAG-NLEDLMSNTNRSINDTGGFMN-LEEDMTRYAYY 118
QY 120 SGIGAGVLVAAYIOVSFWCLAAAGROILKIRKQFFHATMROEIGWEDVDHVGELNTRLTDD 179
DB 119 SGIGAGVLVAAYIOVSFWCLAAAGROIHKIRKQFFHATMROEIGWEDVDHVGELNTRLTDD 178
QY 180 VSKINEGIGDKVGMFFOSFIATFTFTGTVGTPGWKTLVILAIISPVLGLSAAIWAUKLSS 239
DB 179 VSKINEGIGDKVGMFFOSFIATFTFTGTVGTPGWKTLVILAIISPVLGLSAAIWAUKLSS 238
QY 240 FTDKELLAYAKAGAAVEVLAAIRTVIAFGQKKELERYKNLEEKIRIGIKKAITANIS 299
DB 239 FTDKELLAYAKAGAAVEVLAAIRTVIAFGQKKELERYKNLEEKIRIGIKKAITANIS 298
QY 300 IGAAPFLIYASYALAFWYGTSLVLSSEVTIGOVLTVFVSLIGAFSGQASPSIEAFANA 359
DB 299 IGAAPFLIYASYALAFWYGTSLVLSSEVTIGOVLTVFVSLIGAFSGQASPSIEAFANA 358
QY 360 RGAAYEIPKIIDNKPSIDSYSGKHGKPDNKGNEFKNVHFSYPSRKEVKILGKLNKVQ 419
DB 359 RGAAYEIPKIIDNKPSIDSYSGKHGKPDNKGNEFKNVHFSYPSRKEVKILGKLNKVQ 418
QY 420 SGQTVALVNGSGCGKSTTVQLMQRLYDPTDGMVCIQDQRTINVRHLREITGVVSEPV 479
DB 419 SGQTVALVNGSGCGKSTTVQLMQRLYDPTDGMVSVGDQRTINVRHLREITGVVSEPV 478
QY 480 LFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPNKFDTLVGERGASGQKQR 539
DB 479 LFATTIAENIRYGRNVTMDEIEKAVKEANAYDFIMKLPNKFDTLVGERGASGQKQR 538
QY 540 TAIARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVNRNADVI 599
DB 539 TAIARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVNRNADVI 598
QY 600 AGFDDGVIVEKGNHDELMKEGIVFKLVMTQTRNGEIELENATGESKSESDALEMSPKDS 659
DB 599 AGFDDGVIVEKGNHDELMKEGIVFKLVMTQTRNGEIELENATGESKSESDALEMSPKDS 658
QY 660 GSSLIKRRSTRRSIHAPQOQDKLGTCKEDLNENVPVSFWRLKLNSTEWPFYVVGIFCA 719
DB 659 RSSLIKRRSTRRSIHAPQOQDKLGTCKEDLNENVPVSFWRLKLNSTEWPFYVVGIFCA 718
QY 720 IINGGLQPAFSIIFSRIGITRDEDPETKQNSNMFSVLVLVLIISFIFFLQGFTEG 779
DB 719 IINGGLQPAFAIIFSKIIGVETRIDDPETKQNSNLFSLLFLALGIIISFIFFLQGFTEG 778
QY 780 KAGEILTTLKRLVYMFVRSMLRQDVSWFDDPKNTTGALTTLRLANDAAQVKAIGSLAVITQ 839
DB 779 KAGEILTTLKRLVYMFVRSMLRQDVSWFDDPKNTTGALTTLRLANDAAQVKAIGSLAVITQ 838
QY 840 NIANLGTGIIISLIYGWQTLTLLLAIVPIIATAGVVEKMLSGQALKOKKELEGAKTAT 899
DB 839 NIANLGTGIIISLIYGWQTLTLLLAIVPIIATAGVVEKMLSGQALKOKKELEGAKTAT 898
QY 900 EAIENFRTVSTTRQKPFYMAQSLOVPYRNSLRKAHIFGVFSFISITQAMVYFVAGGCFR 959
DB 899 EAIENFRTVSTTRQKPFYMAQSLOVPYRNSLRKAHIFGVFSFISITQAMVYFVAGGCFR 958
QY 960 FGAYLVANEFMFQDVLVFSVIFGAMAVGVSSFPADYAKAKVSAAHVIMIEKSPLI 1019
DB 959 FGAYLVAHKLMFQDVLVFSVIFGAMAVGVSSFPADYAKAKISAHHIIMIEKTPLI 1018
QY 1020 DSYSPHGLKPNTELEGNTVFNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKST 1079

QY 953 SYAGCFREGAYLVANEFMNFODVLVFSVIAVFGAMAVGVQVSFAPDYAKAKVSAHVIMI 1012
Db 955 SYAGCFREGAYLVANEFMNFODVLVFSVIAVFGAMAVGVQVSFAPDYAKAKVSAHVIMI 1014
QY 1013 TEKSLPIDSYPHGLKPNLTGNTVFNENYVPTRPDIPVLQGLSLEVKKGQTLALVGS 1072
Db 1015 TEKTLPIDSYSTEGKLPNTLGNVTNFENVFNPTRLDIPVLQGLSLEVKKGQTLALVGS 1074
QY 1073 SCGKSTVVQLLERYDPLAGSVLIDGKEIKHLNVQWLRHLGIVSQEPILFDCSIAENI 1132
Db 1075 SCGKSTVVQLLERYDPLAGSVLIDGKEIKHLNVQWLRHLGIVSQEPILFDCSIAENI 1134
QY 1133 AYGDNSRVVSHEETMQAAKEANIHHFETLPEKYNTRYGDKGTQLSGGQKQRIARALV 1192
Db 1135 AYGDNSRVVSQEBIVRAAKEANIHAFTESLPNKYSTRVGDKGTQLSGGQKQRIARALV 1194
QY 1193 ROPHTLLDEATSALDTESEKVVQEQALDKAREGRTCVIAHRLSTIONADLIVVFQNGV 1252
Db 1195 ROPHTLLDEATSALDTESEKVVQEQALDKAREGRTCVIAHRLSTIONADLIVVFQNGV 1254
QY 1253 KEHGTQQLLAQKGIYFSMISVQAGAKR 1280
Db 1255 KEHGTQQLLAQKGIYFSMISVQAGAKR 1282

RESULT 12
AA58187
ID AA58187 standard; Protein; 1280 AA.
XX AC AA58187;
XX DT 14-MAR-2000 (first entry)
XX DE Human G185V mutant multidrug resistance-1 (MDR-1) protein.
XX KW Multidrug resistance; MDR-1; P-glycoprotein;
KW transmembrane efflux pump; haematopoietic stem cell; transduction;
KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW gene therapy; gene replacement; genetic defect; thalassaemia;
KW Gauchier's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW cytokine; mutant; mutein.
XX OS Homo sapiens.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 185 /note= "this residue is Gly in the wild-type MDR-1
FT (AA58186)"
FT
FT
XX WO961589-A2.
XX
XX 02-DEC-1999.
XX PD
XX 27-MAY-1999; 99WO-US11825.
XX PF
XX 28-MAY-1998; 98US-0086988.
XX PR
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX PA Sorrentino B, Bunting K;
XX PI WPI; 2000-072615/06.
XX DR N-PSDB; AA249333.
XX
XX Ex vivo expansion of hematopoietic stem cells transduced with a
PT sequence encoding human multidrug resistance-1, used for bone marrow
PT transplantation -
XX
XX Example 1; Page 82-90; 113pp; English.
XX PS
XX This sequence represents human G185V mutant multidrug resistance
CC protein MDR-1, where the Gly residue at position 185

CC of the wild-type protein (AA58186) is replaced by Val. MDR-1 is a
CC transmembrane efflux pump, responsible for the export of drugs from
CC cells, particularly cancer cells. The wild-type MDR-1 shows increased
CC resistance to etoposide and decreased resistance to vinca alkaloids
CC compared with the G185V mutant. The invention relates to transducing
CC haematopoietic stem cells with nucleic acid encoding an MDR protein
CC and culturing the modified cells. The modified haematopoietic stem
CC cells are useful in bone marrow transplantation (to reconstitute
CC haematopoietic systems in patients who have undergone chemotherapy or
CC radiation therapy) and in ex vivo gene therapy of genetic defects in
CC cells derived from haematopoietic stem cells, e.g., thalassaemia,
CC Gauchier's disease, sickle cell anaemia or leukaemia. The modified
CC cells can also be used to identify factors involved in regulating
CC proliferation and differentiation in haematopoietic stem cells.
CC Haematopoietic stem cells that express MDR-1 will be protected against
CC chemotherapeutic agents, so can be engrafted while the patient is
CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC provides sufficient cells to permit standard biochemical analysis.
CC Overexpression of MDR-1 allows cytokine-driven expansion of
CC haematopoietic stem cells by at least 10-fold compared with a maximum
CC of 4-fold in known procedures.
XX Sequence 1280 AA;
XX

Query Match 90.5%; Score 5850.5; DB 21; Length 1280;
Best Local Similarity 90.4%; Pred. No. 0;
Matches 1158; Conservative 59; Mismatches 61; Indels 3; Gaps 3;
QY 1 MDPEGGKKGSA-EKNEFKMGKSKKKEKKTPTVSTFAMFRYSNWLDRLYMLVGTMAALI 59
Db 1 MDLEGRNGGAKKNFKLNKSEKDKKPTVSVFMSFRYSNWLDRLYMLVGTMAALI 60
QY 60 HGAALPLMLLVFGNMTDSFANAGISRNKTPFVILINESITNTQHFHINLHEEEMTYAYY 119
Db 61 HGAGLEPLMLLVFGEMTDIFANAG-NLEDLMSNITNRSINDTGFNMN-LEEDMTRYAYY 118
QY 120 SGIGAGLVAAVYIOVSFCLAGROQLTKRKOFFHAIMRQEGWDFVDHVGELNRLTDD 179
Db 119 SGIGAGLVAAVYIOVSFCLAGROQLTKRKOFFHAIMRQEGWDFVDHVGELNRLTDD 178
QY 180 VSKINEGIDKVGMMFFQSIAITFTGTFIVGTFPGWKLTLVLAISPVLGLSAAIAWAKILSS 239
Db 179 VSKINEVIGDKMGMMFFQSMATEFTGTFIVGTFGRGWLTLVLAISPVLGLSAAVWAKILSS 238
QY 240 FTDKELLAYAKAGAAVEVLAAIRTVIAFGGOKKLEKRNKLEAKRIGIKKAITANIS 299
Db 239 FTDKELLAYAKAGAAVEVLAAIRTVIAFGGOKKLEKRNKLEAKRIGIKKAITANIS 298
QY 300 IGAAFLIIYASALAFWYGTSLVLSSEYITIGOVLTVFFSVLIGAFSIGOAPSIEAFANA 359
Db 299 IGAAFLIIYASALAFWYGTTLVLSGEYSIGOVLTVFFSVLIGAFSIGOAPSIEAFANA 358
QY 360 RGAAYEIFKIIDNKPISIDYSKSGHKPDNIKGNLEFNHVSYPGSRKVKILKGLNLKVQ 419
Db 359 RGAAYEIFKIIDNKPISIDYSKSGHKPDNIKGNLEFNHVSYPGSRKVKILKGLNLKVQ 418
QY 420 SGTVALVGNCGCKSTTVOLMORLYDPTDGMVCIDGQDRTINVRHREITGVVSQEPV 479
Db 419 SGTVALVGNCGCKSTTVOLMORLYDPTDGMVSDGQDRTINVRHREITGVVSQEPV 478
QY 480 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGARSQOKOR 539
Db 479 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGARSQOKOR 538
QY 540 IAIAALVNRPKILLDEATSALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADVI 599
Db 539 IAIAALVNRPKILLDEATSALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADVI 598
QY 600 AGFDDGVIVEKGNHDELMEKGIYFKLVTMOTRGNEIELENATGESKSESDALEMSPKDS 659
Db 599 AGFDDGVIVEKGNHDELMEKGIYFKLVTMOTAGNEIELENADESKSEIDALEMSNDS 658
QY 660 GSSLIKRRSTRRSIIHAPOGQDRKLTGKTEDLNENVPVSFWRLKLNSTNFWVVFVGFCA 719

Db 659 RSSLRKRSTRSRVRSQAQDRKLSTKEALDESIPPVSFWRMKLNLTENPYFVVGVECA 718
 QY 720 IINGLOPAESIIPIESRIIGFTREDDEPETKRONSMESVLFLVGLIISFTIFFLOQFTFG 779
 Db 719 IINGLOPAFAIIPIESRIIGFTREDDEPETKRONSNLFLALGIIISFTIFFLOQFTFG 778
 QY 780 KAGEILTKRLRYMVRSMRLQDVSWFDDPKNTTGALTTRLANDAAOVKGAIGSRSLAVITQ 839
 Db 779 KAGEILTKRLRYMVRSMRLQDVSWFDDPKNTTGALTTRLANDAAOVKGAIGSRSLAVITQ 838
 QY 840 NIANLGTGIIISIIYQWQLPILLIAIIVPIIAIAGVEMKMLSGOALKDKKELEGAGKIAT 899
 Db 839 NIANLGTGIIISIIYQWQLPILLIAIIVPIIAIAGVEMKMLSGOALKDKKELEGAGKIAT 898
 QY 900 EAINENFTVYSLTREQKFEYMAQSLQVYRNSLRKAHIFGVFSFTQAMMYFSYAGCFR 959
 Db 899 EAINENFTVYSLTREQKFEYMAQSLQVYRNSLRKAHIFGVFSFTQAMMYFSYAGCFR 958
 QY 960 FGAYLVANEFMNFQDVLVLSAIVFGAMAYGOVSSFPADYAKAKVSAAHVIMIEKSPLI 1019
 Db 959 FGAYLVANEFMNFQDVLVLSAIVFGAMAYGOVSSFPADYAKAKVSAAHVIMIEKSPLI 1018
 QY 1020 DSYSPHGLKPNTEGNVTFNEVFNPTPRDIPVLOGLSLEVKKGTALVVGSGCGKST 1079
 Db 1019 DSYSTEGMLPNTLEGNNVTFGEVFNPTPRDIPVLOGLSLEVKKGTALVVGSGCGKST 1078
 QY 1080 VVOLLERFYDPLAGSVLIDGKEIKHLNVOLWRAHLGIVSQEPIIFDCSTAEINTAYGDSNR 1139
 Db 1079 VVOLLERFYDPLAGSVLIDGKEIKHLNVOLWRAHLGIVSQEPIIFDCSTAEINTAYGDSNR 1138
 QY 1140 VVSHETIMQAAKANIHHFTIETLPEKYNTRVGDKGTQLSGGQKQRIATARALVRQPHILL 1199
 Db 1139 VVSHETIMQAAKANIHHFTIETLPEKYNTRVGDKGTQLSGGQKQRIATARALVRQPHILL 1198
 QY 1200 LDEATSAIDTESKVVQOEALDKAREGTCIVIAHRLSTIQNADLIYVVFONGKVEHGTQ 1259
 Db 1199 LDEATSAIDTESKVVQOEALDKAREGTCIVIAHRLSTIQNADLIYVVFONGKVEHGTQ 1258
 QY 1260 QLLAQRGIYFMSISVQAGAKR 1280
 Db 1259 QLLAQRGIYFMSISVQAGAKR 1279

RESULT 14
 AAE00306
 ID AAE00306 standard; Protein; 1280 AA.
 AC AAE00306;
 XX AAE00306;
 DT 13-JUN-2001 (first entry)
 DE Human P-glycoprotein (PGP) #1.
 DE Human P-glycoprotein; PGP; multidrug transporter; MDR1;
 KW drug bioavailability; transgenic animal; genetic model.
 XX Homo sapiens.
 OS Homo sapiens.
 PN WO200123540-A2.
 XX 05-APR-2001.
 XX 28-SEP-2000; 2000WO-US26767.
 XX 28-SEP-1999; 99US-0156510.
 XX (GENT-) GENTEST CORP.
 XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 XX WPI; 2001-235373/24.
 XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful

PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors -
 XX Claim 16; Page 78-80; 111pp; English.
 XX The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is human P-glycoprotein (PGP). The
 CC PGP enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.
 XX Sequence 1280 AA;

Query Match 90.5%; Score 5850.5; DB 22; Length 1280;
 Best Local Similarity 90.4%; Pred. No. 0;
 Matches 1158; Conservative 59; Mismatches 61; Indels 3; Gaps 3;

QY 1 MDPEGRKGSA-EKNFMKMGKKKKEKKPTVSTFAMERYSNWLDRLYLVGTMAII 59
 Db 1 MDLEGRNGGAKKNFFKLNKSEKDKKAPTQVSMFYSNWLKLYMVVGTAAII 60
 QY 60 HGAALPLMLVFGNMDSFANAGISRNKTPFVIINESTTNNTTOHFINHEEMTYAYY 119
 Db 61 HGAGLPLMLVFGEMTDIFANAG-NLELMSNITNRSINDTGFNM-LEEDMRYAYY 118
 QY 120 SGIGAGVLVAAYIQVSWFCLAAAGQILKIRKOFFHAIMRQEIFDHDVHGLMTRTDD 179
 Db 119 SGIGAGVLVAAYIQVSWFCLAAAGQILKIRKOFFHAIMRQEIFDHDVHGLMTRTDD 178
 QY 180 VSKINEGIGDKVGMFFQSIATFTFTGIVGTPGKLTIVILAIISPVLGLSAIAWAKILSS 239
 Db 179 VSKINEVIGDKIGMFFQSMATFTFTGIVGTPGKLTIVILAIISPVLGLSAIAWAKILSS 238
 QY 240 FTDKELLAYAKAGAAVEVLAAIIRTVIAFGQKKELERYKNLEAKRIGIKKAITANIS 299
 Db 239 FTDKELLAYAKAGAAVEVLAAIIRTVIAFGQKKELERYKNLEAKRIGIKKAITANIS 298
 QY 300 IGAAPLLIYASYALAFWGTSLVLSSEVTIGOVLTVFESVLIGAFSIGQASPSIEAFANA 359
 Db 299 IGAAPLLIYASYALAFWGTSLVLSSEVTIGOVLTVFESVLIGAFSIGQASPSIEAFANA 358
 QY 360 RGAAYEIFKIIDNKPSIDSYSGKHGKPDNKGLEKFNHVSFSPSRKEVKILKGLNLKVQ 419
 Db 359 RGAAYEIFKIIDNKPSIDSYSGKHGKPDNKGLEKFNHVSFSPSRKEVKILKGLNLKVQ 418
 QY 420 SGQTVLVGNSGCGKSTTVQLMQRLYDPTGMVCIDQDITRTINVRHLREITGVVSOEPV 479
 Db 419 SGQTVLVGNSGCGKSTTVQLMQRLYDPTGMVCIDQDITRTINVRHLREITGVVSOEPV 478
 QY 480 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVCERCARLSGGQKOR 539
 Db 479 LFATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNKFDTLVCERCARLSGGQKOR 538
 QY 540 IAIARALVRNPKILLDDEATSAIDTESAEVQVALDKARKGRITIVIAHRLSVTRADVI 599
 Db 539 IAIARALVRNPKILLDDEATSAIDTESAEVQVALDKARKGRITIVIAHRLSVTRADVI 598
 QY 600 AGFDDGVIVKGNHDELMKEGIFYFLVMTQTRGNEIELENATGESKESDALEMSPKDS 659
 Db 599 AGFDDGVIVKGNHDELMKEGIFYFLVMTQTRGNEIELENATGESKESDALEMSPKDS 658

QY 660 GSSLIKRRSTRSIIHAPOGDRKLTGKEDLNENVPVSFVRILKLNSTWYFVVGIFCA 719
DB 659 RSSLIKRRSTRSVRGSQAQRKLTKEALDESIPPVFWIRLKLNLWYFVVGVECA 718
QY 720 IINGGLQPAFISIFRSIIGIFTRDEDPETKRONSMFVFLVLGIIISIFITFLOGFTFG 779
DB 719 IINGGLQPAFAIFRSKIIGVTRIDDPETKRONSLNLSLLFLALGIIISIFITFLOGFTFG 778
QY 780 KAGEILTKRLRVFRSMRLQRDVDFDDPKNTTGALTTRLANDAAQVKGATGSLRLAVTQ 839
DB 779 KAGEILTKRLRVFRSMRLQRDVDFDDPKNTTGALTTRLANDAAQVKGATGSLRLAVTQ 838
QY 840 NIANLTGTIIISILYGMQLTLLLLAIVPIIAIAGVVENKMLSGQALKDKKKELEGAKIAT 899
DB 839 NIANLTGTIIISIFYGNQTLTLLLLAIVPIIAIAGVVENKMLSGQALKDKKKELEGAKIAT 898
QY 900 EATENFTVVSITREQKFYMAQSLQVYPYRNSLRKAHIFGVSFISITQAMMYFSYAGCFR 959
DB 899 EATENFTVVSITREQKFYMAQSLQVYPYRNSLRKAHIFGVSFISITQAMMYFSYAGCFR 958
QY 960 FGAYLVANEFMFQDVLIVFSAIVFGAMAVGVSSFPADYAKAKYSAHVIMLIEKSPLI 1019
DB 959 FGAYLVANEFMFQDVLIVFSAIVFGAMAVGVSSFPADYAKAKYSAHVIMLIEKSPLI 1018
QY 1020 DSYSPHGLKPNTEGNTVNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKST 1079
DB 1019 DSYSPHGLKPNTEGNTVNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKST 1078
QY 1080 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSTAEINAYGDNRS 1139
DB 1079 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSTAEINAYGDNRS 1138
QY 1140 VWSHEELMQAAKANTHHETLPEKYNTVRGDKGTQLSGGOKORTAJARALVROPHILL 1199
DB 1139 VWSHEELMQAAKANTHHETLPEKYNTVRGDKGTQLSGGOKORTAJARALVROPHILL 1198
QY 1200 LDEATSAIDPESEKVVQEAIDKAREGRTICIVIAHRLSTIONADLIIVFQNGVKYKEHGTQ 1259
DB 1199 LDEATSAIDPESEKVVQEAIDKAREGRTICIVIAHRLSTIONADLIIVFQNGVKYKEHGTQ 1258
QY 1260 QLLAQKGIYFSMVSQAGAKR 1280
DB 1259 QLLAQKGIYFSMVSQAGAKR 1279

RESULT 15
AAP70452
ID AAP70452 standard; Protein; 1280 AA.
XX AAP70452;
XX AAP70452;
XX 21-MAY-1991 (first entry)
XX Sequence encoded by human multi-drug resistance-1 (mdr1) cDNA
XX from clones lambda-HDR10.5 and 104.
XX Chemo-therapy resistant tumour cell; P-glycoprotein.
XX Homo sapiens.
XX W08705943-A.
XX 08-OCT-1987.
XX 26-MAR-1987; 87WO-US00758.
XX 01-AUG-1986; 86US-0892575.
XX 28-MAR-1986; 86US-0845610.
XX (UNII) UNIV OF ILLINOIS.
XX Roninson IB, Pastan IH, Gottesman MM;
XX

DR WPI; 1987-291656/41.
XX N-PSDB; AAN70752.
XX DNA for multi-drug resistance in human cells - used to detect
PT chemotherapy-resistant tumour cells and for producing
PT polypeptide(s) for diagnosis and therapy
XX Claim 4(a); Table 5, pp30-39; 61pp; English.
XX The human multi-drug resistant KB carcinoma cell lines were used as
CC the source of the mdr1 gene nucleic acid sequences (AAN70751). To
CC obtain cDNA clones of the mdr1 gene (AAN70752), poly (A) and RNA was
CC used. Analysis of the AA sequence presented in (AAN70752) indicates
CC that the mdr1 gene product is likely to be a transmembrane protein.
CC The presence of transmembrane domains and potential glycosylation
CC sites is consistent with the mdr1 protein being related to the
CC P-glycoprotein.
XX
XX
SQ Sequence 1280 AA;
Query Match 90.4%; Score 5847.5; DB 8; Length 1280;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 1157; Conservative 60; Mismatches 61; Indels 3; Gaps 3;
QY 1 MDPGGRKGS-A-EKNFWKMGKSKKKEKPTVTSTAMPRYSNWLDRMLYGLTMAAII 59
DB 1 MDLEGRNGGAKKNEFKLNKSEKDKKPTVSFMSFRYSNWLDRMLYGLTMAAII 60
QY 60 HGAALPLMLVFGNMTDSFANAGISRNTFPVILINESITNTNTOHFINHLEEMTYAYY 119
DB 61 HGAGLPLMLVFGEMTDIFANAG-NLEDMSNITNRSNDINDTGFFMN-LEEDMTRYAYY 118
QY 120 SGIGAGVLVAAYIQVSWCLAAAGRIKIRKQFFHAIMRQEIWFVDVHVGELNTRLTD 179
DB 119 SGIGAGVLVAAYIQVSWCLAAAGRIKIRKQFFHAIMRQEIWFVDVHVGELNTRLTD 178
QY 180 VSKINEGIDKVGMPFQSIATFTFCIVGFTPGWKLTVILAIAPVLGSLAAIWAKILSS 239
DB 179 VSKINQVIGDKIGMPFQSIATFTFCIVGFTPGWKLTVILAIAPVLGSLAAIWAKILSS 238
QY 240 FTDKELLAYAKAGAAVEVLAAIRTVIAFGGKKELERYKNLEAKRIGIKKAITANIS 299
DB 239 FTDKELLAYAKAGAAVEVLAAIRTVIAFGGKKELERYKNLEAKRIGIKKAITANIS 298
QY 300 IGAAFLLIYASALAFWYGTSLVLSSEYTIQVLTVPFSLVIGAFSIGQASPSIEAFANA 359
DB 299 IGAAFLLIYASALAFWYGTSLVLSSEYTIQVLTVPFSLVIGAFSIGQASPSIEAFANA 358
QY 360 RGAAYEIKIDNKPISIDSKSGHKPDNIKGNLFKFNHFSYPSRKEVKILGLNLKVQ 419
DB 359 RGAAYEIKIDNKPISIDSKSGHKPDNIKGNLFKFNHFSYPSRKEVKILGLNLKVQ 418
QY 420 SQGTVALVNGSGCKSTTVQLMQRLYDPTDGMVCDIGDQIRTIINVRHLREITGVVSQEPV 479
DB 419 SQGTVALVNGSGCKSTTVQLMQRLYDPTDGMVCDIGDQIRTIINVRHLREITGVVSQEPV 478
QY 480 LPATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPKPHFDILVGERGQLSGGQKOR 539
DB 479 LPATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPKPHFDILVGERGQLSGGQKOR 538
QY 540 IAIARALVRNPKILLDEATSAIDPESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 599
DB 539 IAIARALVRNPKILLDEATSAIDPESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI 598
QY 600 AGFDDGVIVEKGNHDELMKEGIYFKLVTMQTRGNEIELENATGSKESDALEMSPKDS 659
DB 599 AGFDDGVIVEKGNHDELMKEGIYFKLVTMQTRGNEIELENATGSKESDALEMSPKDS 658
QY 660 GSSLIKRRSTRSIIHAPOGDRKLTGKEDLNENVPVSFVRILKLNSTWYFVVGIFCA 719
DB 659 RSSLIKRRSTRSVRGSQAQRKLTKEALDESIPPVFWIRLKLNLWYFVVGVECA 718
QY 720 IINGGLQPAFISIFRSIIGIFTRDEDPETKRONSMFVFLVLGIIISIFITFLOGFTFG 779
DB 719 IINGGLQPAFAIFRSKIIGVTRIDDPETKRONSLNLSLLFLALGIIISIFITFLOGFTFG 778

Db 719 IINGLOPAFAIIFSIIIGVETRIDDPETKQNSNLSLFLALGIISFITFFLQGTFG 778
Qy 780 KAGELLTKRLRYMYFRSMLRODVSWFDDPKNTTGALTTRLANDAAQVKGAGSLAVITQ 839
Db 779 KAGELLTKRLRYMYFRSMLRODVSWFDDPKNTTGALTTRLANDAAQVKGAGSLAVITQ 838
Qy 840 NIANLGTGIIISLIYGWLTLLLAIVPIIAIAGVVEKMLSGOALKDKKELEGAGKIAT 899
Db 839 NIANLGTGIIISFIYGWLTLLLAIVPIIAIAGVVEKMLSGOALKDKKELEGAGKIAT 898
Qy 900 EAIENFRTVSLTREKPEYMAQSLQVPYRNSLRKAHIFGVSPSITQAMMYFSYAGCFR 959
Db 899 EAIENFRTVSLTQEQFEHMYAQSLQVPYRNSLRKAHIFGITFTQAMMYFSYAGCFR 958
Qy 960 FGAYLVANEFMNFODVLLVFSAYIFGAMAYQVSSFPADYAKAKVSAHVIMIEKSPLI 1019
Db 959 FGAYLVAKLMSFEDVLLVFSAYVFGAMAYQVSSFPADYAKAKISAHHIIMIEKTPLI 1018
Qy 1020 DSYSPHGLKPNLTLEGNVTFNEVFNYPTRDIPVLOGLSLEVKKGOTLALVSGSGCKST 1079
Db 1019 DSYSTEGMLMNTLLEGNVTFGEVFNYPTRDIPVLOGLSLEVKKGOTLALVSGSGCKST 1078
Qy 1080 VVOLLERFYDPLAGSVLIDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAENIAYGDNRSR 1139
Db 1079 VVOLLERFYDPLAGKVLIDGKEIKRLNVQWMLRAHLGIVSQEPILFDCSIAENIAYGDNRSR 1138
Qy 1140 VWSHEETMOAAKEANIIHFIETLPEKYNTRVGDKGTOLSGGQKQRTAIARALVRQPHILL 1199
Db 1139 VVSQEEIVRAAKEANIIHAFIESLPNKYSTKVGDGTQLSGGQKQRTAIARALVRQPHILL 1198
Qy 1200 LDEATSALDTESEKVVQEQALDKAREGRTCIVIAHRLSTIONADLIIVVFQNGKVEHGTQ 1259
Db 1199 LDEATSALDTESEKVVQEQALDKAREGRTCIVIAHRLSTIONADLIIVVFQNGKVEHGTQ 1258
Qy 1260 QLLAQKGIYFSMISVQAGAKR 1280
Db 1259 QLLAQKGIYFSMVSVQAGTKR 1279

Search completed: November 6, 2002, 18:39:15
Job time : 39.4697 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:45:15 : Search time 3800.84 Seconds
(without alignments)
7047.385 Million cell updates/sec

Title: us-09-672-725c-4

Perfect score: 6467

Sequence: 1 MDEGGRKGAENKFWKMGK.....LLAQKGIYFSMISYQAGAKR 1280

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/Cgn2.1/USPTO.scool/US09672725/runat_04112002_124341_17977/app_query.fasta_1.6435
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725_eCgn.1.11409_eRunat_04112002_124341_17977 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htgo.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	6467	100.0	4317	4	AF045016 Canis fam
2	6467	100.0	4317	6	AX105059 Sequence
3	6428.5	99.4	4045	12	AF269224 Synthetic
4	6425.5	99.4	4279	6	AX105078 Sequence
5	6425.5	99.4	4279	6	AX105080 Sequence
6	6425.5	99.4	4279	6	AX105082 Sequence
7	6420.5	99.3	4279	6	AX105057 Sequence
8	6385	98.7	3934	4	CF419568 Canis fam
9	5859.5	90.6	3860	6	AX322787 Sequence
10	5856.5	90.6	4378	6	E02326 Multidrug r
11	5855.5	90.5	4186	6	AX108654 Sequence
12	5853	90.5	4195	6	AX108656 Sequence
13	5850.5	90.5	3860	6	AX322789 Sequence
14	5850.5	90.5	4646	6	AX336420 Sequence
15	5850.5	90.5	4646	6	AX336708 Sequence
16	5850.5	90.5	4646	6	I49610 Sequence 2
17	5850.5	90.5	4646	9	HUMMDR1
18	5850.5	90.5	4669	6	I08557 Sequence 3
19	5850.5	90.5	6505	6	AR028671 Sequence
20	5850.5	90.5	8630	6	AX012320 Sequence
21	5850.5	90.5	8630	6	AX012321 Sequence
22	5850.5	90.5	9318	6	AR028672 Sequence
23	5843	90.4	3988	6	AX024454 Sequence
24	5843	90.4	4192	9	AF016535 Homo sapi
25	5823.5	90.0	4669	6	AR091275 Sequence
26	5809	89.8	4264	6	AR051647 Sequence
27	5809	89.8	4264	6	AR051650 Sequence
28	5790.5	89.5	4669	6	AR055785 Sequence
29	5666	87.6	3858	4	OU78609 Ovis aries
30	5661	87.5	4296	10	CRUPEPI
31	5650	87.4	4304	10	CRUPEPI165
32	5638.5	87.2	4788	6	AX322793 Sequence
33	5638.5	87.2	4924	10	MUSMDR1A
34	5634.5	87.1	4927	10	AF257746 Rattus no
35	5631.5	87.1	4356	10	MUSMDR1A
36	5612.5	86.8	4323	10	AF286167 Rattus no
37	5505.5	85.1	3489	4	AB029153 Felis cat
38	5389	83.3	3987	10	CRUPEPI185
39	5304.5	82.0	4189	6	AX322791 Sequence
40	5304.5	82.0	4298	10	MUSMDR
41	5296	81.9	4233	6	AR123273 Sequence
42	5281.5	81.7	4280	10	CRUPEPII
43	5236	81.0	4254	10	RATMDRM
44	4982	77.0	3682	10	MUSPL1A
45	4898	75.7	3924	6	AX024455 Sequence

ALIGNMENTS

RESULT 1
AF045016
LOCUS AF045016
DEFINITION Canis familiaris multidrug resistance p-glycoprotein (MDR1) mRNA,
complete cds.
ACCESSION AF045016
VERSION AF045016.1 GI:2852440
KEYWORDS
SOURCE dog.

```

ORGANISM  Canis familiaris
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
TITLE      1 (bases 1 to 4317)
JOURNAL    Puel, O., Lepage, J.F., Alvinerie, M., Galtier, P. and Pineau, T.
           Direct Submission
           Submitted (28-JAN-1998) Pharmacology, INRA, BP 3, 180 Chemin de
           Tournereuil, Toulouse Cedex 9 31931, France
FEATURES   Location/Qualifiers
            1..4317
             /organism="Canis familiaris"
             /db_xref="taxon:9615"
             /tissue_type="colon"
             /note="beagle"
            1..4317
             /gene="MDR1"
            70..3912
             /gene="MDR1"
             /codon_start=1
             /product="multidrug resistance p-glycoprotein"
             /protein_id="AAC02113.1"
             /db_xref="GI:2852441"
             /translation="MDPEGGRKSAENFWMKGGKSKKKKPTVSTFAMFYSNW
LDRLVLMVGTMAAIIHGAALPLMLVFNMTDSFANAGISRNKTFPVIINESTINNTQ
HFINLEEMTYAYVYSGIGAGLVAAVIOVSFWCLAAAGRIKIRKQFFHLMROE
IGWPFVHDVGLNLTLDVSKINEGIGDKVGFQSIATFTFTGFIIVFTGPKMLTIV
ILAI5PVLGSNAIHWAKILSFTDKELLAYAKAGAAVEVLAAIIRTIVFGGKKLE
RYKNLEAKRIGIKKAITANISIGAALFLLIYSAALAFWGTSLVLSSEYTGQVLT
VFFSLIGAFSFGQASPTSEFANARGAAEYFKLIIDNKPSIDISKSGHAPDNKGN
LEFNWHSYSPSRKEVKILKGLNLKVSQGVVALVNGSCGKSTVQLMRLYDPTDG
MYCIDQDIRTINVRHLREITGVVSEQVLPFATTAENIRYGRNVTMDTEKAVKA
NAYDFILPNKDFLVRGARLGGOKRIATARALVRNPKILLDEATSAIDTES
EAVQVADKABKGTIVIAHRLSTVRNADVIAGDDGVIVEKGNHSDHLMKEKGIVF
KLVTKQDNGNELENAATGESKSESDALEMPKDSGLIKRRSTRRSIIFSRILGIF
LGTDEPETKQNSMPSVFLVILGIIISFTFFLQGTFTFGKAGEILITKRLRIVFRSML
RODVSFDDPNKNTGALTRLNDAQVKAIGSLAVITONIANLGTIIISLYGW
QTLULLALVPIIATAGVEMKMLSGQALKDKKELEGAKIATEAENFRVWSLIRE
OKFEYMAQSLQVYRNSLRKAHIFGVFSITQAMMYFSYAGCFRFGAYLVANEFMNF
ODVLLVFSIAVFGMAAGVSSFPADYAKAKVSAHVIMITEKSLIDISYSPHGLKPN
TLEGNTFNEVFNPTRPDIPLVGLSLIEVKKGTALVGGSSCGKSTVQLQLERFY
DPLAGSVLIDGKEIKHLNQLMRLAHLGIYSQEPILFDCSIAENIAYGNSRVSHET
MQAAKEANTHHPIETLPEKYNTRVGDGKQQLSGOKQRIATARALVRPHILLDEAT
SALDTESESRVQVQALDKAREGRTCTIVIAHRUSTIONADLIVVFQNGKVEHGTQQLL
AOKGIYFSMISVQAGAKR"
BASE COUNT 1293 a 844 c 1019 g 1161 t
ORIGIN
Alignment Scores:
pred. No.: 0 Length: 4317
Score: 6467.00 Matches: 1280
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-672-725C-4 (1-1280) x AF045016 (1-4317)
Qy 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 70 ATGGATCTCTGAAGGAGCGGTGAAGGGAGTGTCAGAGAAAGAACTCTGGAATAATGGGCAAA 129
Qy 21 LysSerLysLysGluLysLysGluLysLysProThrValSerPheAlaMetPheArg 40
Db 130 AAAAGTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 189
Qy 41 TyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleHis 60
Db 190 TATTCAAAATGGCTGTATGTATGTGTGTGGGGAAGTGGCTGCCATCATCCAT 249
Qy 61 GlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAsn 80
Db 250 GGAGCTGCACACCCCTCTCATGATGCTGGTTTGGAAACATGACAGATAGCTTTGGCAAAAT 309

```

```

Qy 81 AlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsn 100
Db 310 GCAGGAATTTCAAGAAACAAACTTTTCCAGTTATAATTAATGAAAGTATTACGAACAAT 369
Qy 101 ThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyrSer 120
Db 370 ACACAAATTTTCATCAACCATCTGGAGGAGGAATGACCACGTATGCTTATTTATTCAGT 429
Qy 121 GlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeuAla 140
Db 430 GGGATCGGTCTGGCGTCTGGTGGCTGTATCATCCAGGTTTCATTCGTGCTCGCTGGCA 489
Qy 141 AlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGlu 160
Db 490 GCAGGAAGACAGATACTCAAAATTAGAAAACAATTTTTCATGCTATCATGCGACAGGAG 549
Qy 161 IleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAspVal 180
Db 550 ATTGGCTGGTTGACGTGCATGAGCTTGGGAGCTTAACACCCGGCTCACAGCATGTC 609
Qy 181 SerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAlaThr 200
Db 610 TCCAAATCAATGAAGAAATGGCGACAAAGTTGGAATGTTCTTCAATCAATAGCAACA 669
Qy 201 PhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIleLeu 220
Db 670 TTTTTCACCGGTTTATAGTGGGCTTTACACCTGTTTGGAAAGCTTAACCCCTGTGATTTG 729
Qy 221 AlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPhe 240
Db 730 GCCATACGCCCTGTCTTGGACTTTCAGCCGCCATCTGGGCAAGATACATATCTTCATT 789
Qy 241 ThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeuAla 260
Db 790 ACTGATAAAGAACTCTTGGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGCTTAGCA 849
Qy 261 AlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsnLys 280
Db 850 CCAATCAAGAACTGTATGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACACAAA 909
Qy 281 AsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSerIle 300
Db 910 RATTAGAAAGAACTTAAAGAAATTTGGGATAAAGAAAGCTATCAGCGGCAACATTTCTATT 969
Qy 301 GlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSer 320
Db 970 GGTGCGGCTTCTTATTGATCATGATCATGCTCTGGCTTTCTGTTATGGACCTCC 1029
Qy 321 LeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerValLeu 340
Db 1030 TTGGTCTCTCCAGTGAATATATATCTATTGGACAGGTACTCACTGCTCTTCTTCTGTATTA 1089
Qy 341 IleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArg 360
Db 1090 ATTTGGGCTTTTATGATTGGACAGCATCCCAAGCATTTGAAGCATTTGCAAAACGCAAGA 1149
Qy 361 GlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSer 380
Db 1150 GGAGCAGCTTATGAATCTTCAAGATAATTTGACATAATTAACCAAGCATTTGACAGCTATTCCG 1209
Qy 381 LysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPhe 400
Db 1210 AGAGTGGACATAAACACAGATAATATTAAAGGAAATTTGGAATTTCAAAATGTTCACTTC 1269
Qy 401 SerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSer 420
Db 1270 AGTTACCTTCTCGAAAAGAAAGTAAAGATCTTAAAGGGGTCTCAACCTGAAGGTTCCAGAGT 1329
Qy 421 GlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeu 440
Db 1330 GGGCAGACAGTGGCGCTGTTGGGAACAGTGGCTCGGGGACAGCACGACCGTCAGCTG 1389

```

QY	441	MetGlnArgLeuTyrAspProThrSerPglyMetValCysIleAspGlycInAspileArg	460
DB	1390	ATGCAGAGGCTCTATGACCCACAGATGGCATGCTGTATTGATGGACAGACATTAGG	1449
QY	461	ThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeu	480
DB	1450	ACCATAAATGTAAAGCATCTTCGGGAATTTACTGGTGTTGGTGAGTCAGAGCGCTGTGTG	1509
QY	481	PheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAspGlu	500
DB	1510	TTTGCCACCACGATAGCTGAAAACATTCGCTATGCGCGCGGAAAAATGCACCATGGATGAG	1569
QY	501	IleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMeLysLeuProAsnLys	520
DB	1570	ATTGAGAAGCTGTTAAGGAAGCAATGCTATGATTTTTATCATCAAACCTACCTAATAAAA	1629
QY	521	PheAspThrLeuValGlyLeuArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArgIle	540
DB	1630	TTTGACACTCTGGTTGGAGAGAGGGGCCGCTGAGTGGTGGACAGAAAACAGAGAATC	1689
QY	541	AlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThrSer	560
DB	1690	GCCATTGCTGGGCCCTGGTTGCCAACCCCAGATCTCTCTGCTGGATGAGCAACGTCAT	1749
QY	561	AlaLeuaspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLysGly	580
DB	1750	GCTCTGGACACTGAAAGTAGACAGTGGTTTCAGGTGGCCTGGATTAAGCCAGAAAAAGC	1809
QY	581	ArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla	600
DB	1810	CGGACTACCATTTGGATGAGCTCATCGTTGTCTACAGTTCGTAAATGCCGATGTCAATGCT	1869
QY	601	GlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGluLys	620
DB	1870	GGTTTTCATGATGGAGTCATTGTGGAGAAGAAGATCATGATGAACATCATGAAGAAGAG	1929
QY	621	GlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGluAsn	640
DB	1930	GGCATTTTACTTCAAACTTGTCAACAATGCAGACAAGAGAAATGAAATGGATTAGAAAT	1989
QY	641	AlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSerGly	660
DB	1990	GCCACTGGTGAATCCAAAAGTGAAGTGAATGCTTGGAAATGTCTCCAAAAGATTCAGGG	2049
QY	661	SerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAsp	680
DB	2050	TCCAGTTTATAAAAAGAGATCAACTCCAGGAGTATACATGCACCAAGGCCAAGAC	2109
QY	681	ArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrpArg	700
DB	2110	AGAAAGCTGGTACAAAAGAGGACTTGAATGAGAATGTTCTTCAGTTTCTCTCTGGAGG	2169
QY	701	IleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAlaIle	720
DB	2170	ATTCTGAAGCTGAACCTCAACTGAATGGCCTATTATTGGTGGTGATATTTTGTGCTATT	2229
QY	721	IleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePhe	740
DB	2230	ATAAACGGAGCGCTGCAACACGACATTTCAATATATATTTTCAAGGATATATAGGGATCTT	2289
QY	741	ThrArgAspGluaspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe	760
DB	2290	ACCCGAGATGAGGATCCTGAAACAAAACAGACAGAAATAGTAACATGTTTTCTGTATTGTTT	2349
QY	761	LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys	780
DB	2350	CTAGTCTCTGGAAATATTTCCTTTTATTACATTTTTCTCCAGGGCTTCACATTTGGCAAA	2409
QY	781	AlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGln	800
DB	2410	GCTGGGGAGATCCTCACATAAGCGGCTCGATACATAGTGTTCAGATCCATGCTGAGACAG	2469
QY	801	AspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla	820

Db	2470		GATGTCAGCTGGTTTGATGACCCATAAAACACCACCTGGAGCATTGACACACAGGCTTGCC	2529
QY	821		AsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn	840
Db	2530		AATGATGGCGGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTCATTAACCCAGAAT	2589
QY	841		IleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeu	860
Db	2590		ATAGCAATCTTGGACAGGCATATTATATCTTAACTATGTTGGCAATTACACTT	2649
QY	861		LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMetLeu	880
Db	2650		TTACTTCTAGCAATTGTACCATCATTCGCAATAGCAGGAGTTGTTGAATGAAATGTTG	2709
QY	881		SerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGlu	900
Db	2710		TC TGACACAGCAGCTGAAGATAGAAGAGCTAGAAGAGAGCTGGGAAGATTGCTACAGAA	2769
QY	901		AlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyrMet	920
Db	2770		GCATCGAAAACCTCCGAACTGTTGTTCTTTGACTCGGAGCAGAAGTTTGAATACATG	2829
QY	921		TyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGly	940
Db	2830		TATGCACAGAGTTTGCAGTACCATACAGAAACTCTTTCGAGGAAGACACATCTCGG	2889
QY	941		ValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPhe	960
Db	2890		GTCTCATTTTCTATCACCCAGGCAATGATGATTATTTTCTATGCTGGCTGTTTCCGGT	2949
QY	961		GlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSer	980
Db	2950		GTGGCTACTTGTGGGCAATGAGTTCTATGAACTTTCAGGATGTCTTTTGGTATTCTCA	3009
QY	981		AlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyrAla	1000
Db	3010		GCTATTGTCTTGTGGCCATGGCGAGTGGCGAGTCAGTTCACTTTCCTGCTACTATGCC	3069
QY	1001		LysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIleAsp	1020
Db	3070		AAAGCCAAAGTATCAGCAGCCACGCTCATCATGATCATTTGAAAAGGCCCTCTGATTGAC	3129
QY	1021		SerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGlu	1040
Db	3130		AGCTACAGCCCTCAGGCCCTCAGGCCAAATACGTTGGAAAGAAATGTACATTTAATGAG	3189
QY	1041		ValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeuGlu	1060
Db	3190		GTCGTGTCAACTATCCACTCGACACAGACATCCCGCTGCTCCAGGGGCTGAGGCTCGAG	3249
QY	1061		VallysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrVal	1080
Db	3250		GTGAAGAAGGGCCAGACGCTGGCCCTCGTGTAGTAGCAGTGGCTGTGGGAAGAGCACAGT	3309
QY	1081		ValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGlyLys	1100
Db	3310		GTTCAGCTCCTAGAGGCTTCTATGACCCTTGGCTGGTTCTCAGTGTCAATTGATGGCAA	3369
QY	1101		GluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGlu	1120
Db	3370		GAGATAAGCACCTGAATGTCCAGTGGCTCCGAGCACACCTGGGCATCGTCTCAGGAG	3429
QY	1121		ProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgVal	1140
Db	3430		CCATCTCTGTTTACTGTCAGCATTCGCCAGAACATTCCTATGAGACAACACGCCGGTCTC	3489
QY	1141		ValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisPheIleGlu	1160
Db	3490		GTATCATGAAGAGATTATGCAGGAGCCGAAGAGGGCCAAATACACCACTTCATCTCAG	3549
QY	1161		ThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGly	1180

```
Db 3550 ACACCTCCCTGAGAAATACAAACACACAGACTAGGAGAGAAAGGAAACCCAGCTCTCTGGTGGC 3609
Qy 1181 GlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeu 1200
Db 3610 CAGAAACAGCGCATGGCCATAGCTCGCGCTCTGTGTAGACAGCCATATTTTGGCTTTG 3669
Qy 1201 AspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGlnAlaLeuAsp 1220
Db 3670 GATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAGAAGCCCTGGAC 3729
Qy 1221 LysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsn 1240
Db 3730 AAAGCCAGAGAAAGCGCACCTGCATTTGTATCGCCACCGCTTGTCCACCATCCAGAAT 3789
Qy 1241 AlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHisGlnGln 1260
Db 3790 GCAGATTTTAATAGTGTGTTCAGAAATGGCAAAAGTCAAGGAGCATGGCCACATCAACAG 3849
Qy 1261 LeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLysArg 1280
Db 3850 CTGCTGGCTCAGAAAGGCATCATTTTCCATGATCAGTGTCCAGGCTGGAGCAAGCGC 3909

RESULT 2
AX105059 4317 bp DNA linear PAT 30-APR-2001
LOCUS Sequence 3 from Patent WO0123540.
DEFINITION AX105059
ACCESSION AX105059
VERSION AX105059.1 GI:13921211
KEYWORDS dog.
SOURCE
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
Stocker,P.J., Steimel-Crespi,D.T., Crespi,C.L., Reif,T.C. and
Patten,C.J.
P-glycoproteins and uses thereof
Patent: WO 0123540-A 3 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
source
1..4317
/organism="Canis familiaris"
/db_xref="taxon:9615"
70..3912
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37719.1"
/db_xref="GI:13921212"
/translation="MDPEGRKGSAEKNFWMKSKSKEKKKPTVSTFAMFRYSNW
LDRLYLVGTMAAIITHGAALPLMLVFGNMTDSFANAGISRNKTFPVIINESITNNTQ
HPINLHEEMTYAYYSIGAGVLAAYIQVSEWCLAAAGROILKIRQFFHAIHQE
IGWFDVHDVGLNRLTDDSVKINEGIDGKVMFFQSIATFTFTGIVGTFPGWKLTIV
ILAISPVLGSAIAWAKILSSPTDKELIAYAKAGAAVEVLAAIRTVIAFGGKKELE
RNNKLEENKRGIKATANTISIGAFLLIYASVALAFWGTSLVLSSEYTIQGVLT
VFFSVLISIGQASPSIEAPANARGAAVEILFIIDNKPISIDTSKSGHDPDIKGN
LEFNKVFHSPYRKEVKILKLNLVQSGQTVLVNCGSKSTTVQLMRLYDPTDG
MYCIDQDITINVRHLREITGVWQEPVLFATTAENIRYGRNVMTDETEKAVKEA
NAYDFTLKPNEFDLTVGERKAGSGQKQRIAIARALVRNPKILLDEATSAIDTES
EAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGVIKEGNHDELMEKRGYIF
KLYVTQGRNEFSELENAKESSEDALEMPKDSGSSILKRRSPRRHSIAPQOGDRK
LGTKEDLNESFVSFRILKUNSTENFYVVGIFRCALINGLQPAFSIIFSRIGIF
TRDEDPETKRNSMFSVFLVLGIIISIFITPFGFTFGKAGEIILTKRLRTWFRSML
RQDSWFDPPKNTTGALTTLANDAAVKAGSRLAVITQNIANLNGTIIISLIYWG
QUTLLIILAIPTIATVAGVEMKHLGQALDKKKELEGAGKATEAENFRVWLSITRE
OKFEYMAQSLQVYRNSLRKAHIFGVFSFTIQAMMYEYAGCFRFGAYLVANEFMNF
QDVLVFSIIVGMAVAGVSSFPADYAKAVKSAHVIMIEKSLPIDSYPHGLKPN
TLEGNVTFNEVYFNPTRDPIVLOGLSLEVKKGQTLALVSGSGCGKSTVQLLRFY
DPLAGSLVDGKEIKHLNVLQWLRHLGIVSQEPIFLDCSIAENIAYGNSRVVSHKEI
MQRAKEANTHHFETLPEKYNTRVGDKRGQKQRIATARALVRNPKHILLDEAT
SALDTESEKRVQEQALDKAREGRTCIVIAHRLSTIQNALDLIVVFQNGKVKHEGTHQQLL
AQKGIYFSMISVQAGAKR"
1293 a 844 c 1019 g 1161 t

BASE COUNT
ORIGIN
```

```
Alignment Scores:
Pred. No.: 0 Length: 4317
Score: 6467.00 Matches: 1280
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-672-725C-4 (1-1280) x AX105059 (1-4317)
Qy 1 MetaspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 70 ATGGATCCTCAAGAGAGCCGTAAGGGGAGTGCAGAGAAAGAACTTCGGAAATATGGCAAA 129
Qy 21 LysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPheArg 40
Db 130 AAAGTAAAAAGAGAGAAGAAAGAAACCAACTGTCAGACAGCTTGTGCAATGTTTCGC 189
Qy 41 TyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleHis 60
Db 190 TATTCAAAATGGCTTGATAGCTTGTATATGTTGGTGGGACAATGGCTGCATCATCCAT 249
Qy 61 GlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAsn 80
Db 250 GGAGCTGCACCTCCCTCTCATGATGCTGCTGTTTGGAAACATGACAGATAGCTTGCAAAT 309
Qy 81 AlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsn 100
Db 310 GCAGGAATTTCAAGAAACAAAACTTTTCCAGTTATAATAATGAAGTATTACGAACAAT 369
Qy 101 ThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyrSer 120
Db 370 ACACAACATTTTCATCAACCATCTCGAGGAGAAATGACACGATGCTATTAATACAGT 429
Qy 121 GlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeuAla 140
Db 430 GGGATCGGTGCTGGCGTGTGCTGCTGTACATCAGGTTTCATCTGGTGGCTGGCA 489
Qy 141 AlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGlu 160
Db 490 GCAGGAAGACAGATACTCAAAATTTAGAAAACAATTTTTCATGCTATCATCGCAGAGAG 549
Qy 161 IleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAspVal 180
Db 550 ATTGGCTGGTTGACGTGATGACGTGGGAGCTTAAACCCCGGCTACAGACCATGTC 609
Qy 181 SerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAlaThr 200
Db 610 TCCAAATCAATGAAGAAATGGCGACAAAGTTGGAATGTTCTTTCAATCAATAGCAACA 669
Qy 201 PhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIleLeu 220
Db 670 TTTTTCACCGGTTTTATAGTGGGTTTTACACTGTTGGAAAGCTTAACCTTGTGATTTTG 729
Qy 221 AlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPhe 240
Db 730 GCCATCAGCCCTGTTCTTGGACTTTTACGCCCATCTGGGCAAGAGATACATCTCATTT 789
Qy 241 ThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluLeuAla 260
Db 790 ACTGATAAAGAACTCTTGGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTAGCA 849
Qy 261 AlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluAtrqTyrAsnLys 280
Db 850 GCAATCAGAACTGTGATTTGCCCTTTGGAGACAAAAGAAAGAACTTGAAGGTACACAAA 909
Qy 281 AsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSerIle 300
Db 910 AATTTAGAAGAAAGCTTAAAGAAATGGGATAAAGAAAGCTATCACGGCAACATTCATT 969
Qy 301 GlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSer 320
```


Db 970 GGTGGCGCTTCTTATTGATCATATGCTCTGGCTTCTGGTATGGACCTCC 1029
Qy 321 LeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerValLeu 340
Db 1030 TTGGTCCCTCCAGTGAATATACATATTGGACAGGTACTCACTGCTCTCTTTCTGTATTA 1089
Qy 341 IleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArg 360
Db 1090 ATTGGGCTTTAGTATTGGACAGCATCCCAAGCATTTGAAGCATTTGCAACGCCAAGA 1149
Qy 361 GlyAlaAlaTyrGluIlePheLysIleLeuAspAsnLysProSerIleLeuAspSerTyrSer 380
Db 1150 GGACAGCTTATGAATCTTCAAGATAATTGACAAATAAACCAAGCATTTGACAGTATTCG 1209
Qy 381 LysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPhe 400
Db 1210 AAGAGTGGACATAAACCAAGATAATTAAAGGGAATTTGGAATTCAAAAATGTTCACTTC 1269
Qy 401 SerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSer 420
Db 1270 AGTTACCCTTCTCGAAAGAAGTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTACAGT 1329
Qy 421 GlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeu 440
Db 1330 GGGCAGACAGTGGCGCTGGTTGGGACAGTGGCTGCGGGAAGACACGACCGTGCAGCTG 1389
Qy 441 MetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArg 460
Db 1390 ATGCAGAGGCTCTATGACCCACAGATGGCATGGCTGTGTATTGATGGACAGGACATTAGG 1449
Qy 461 ThrIleAsnValArgHisLeuValGluIleThrGlyValValSerGlnGlnProValLeu 480
Db 1450 ACCATAAATTAAGGCATCTTCGGAAATTAAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1509
Qy 481 PheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAspGlu 500
Db 1510 TTTGCCACCAAGATAGCTGAACACATTCGCTATGGCGCGAAATGTCCACATGGATGAG 1569
Qy 501 IleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsnLys 520
Db 1570 ATTGAGAAGCTGTTAAGGAAGCAATGCCTATGATTTTATCATGAACATCACTTAATAAA 1629
Qy 521 PheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArgIle 540
Db 1630 TTTGACACTCTGTGTGGAGAGAGGGCGCGCTGAGTGGTGACAGAAACACAGAGAATC 1689
Qy 541 AlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThrSer 560
Db 1690 GCCATTGCTGGGCGCTGGTTCCGAACCCCAAGATTCTTCTGCTGGATGAGGCAAGTCA 1749
Qy 561 AlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLysGly 580
Db 1750 GCTCTGGACACTGAAGAGTGAAGCAGTGGTTCAAGTGGCCCTGGATAAGGCCAGAAAGGC 1809
Qy 581 ArgThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla 600
Db 1810 CGGACTACCATTTGTATAGCTCATCTGTTGTCTACAGTTTCGTAATGCCGATGTCATTGCT 1869
Qy 601 GlyPheAspGlyValIleValGlyGlyAsnHisAspGluLeuMetLysGluLys 620
Db 1870 GGTGTTGATGGAGTCATTTGGGAAGGAATCATGATGAATCTCATGAAGAGAGAG 1929
Qy 621 GlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGluAsn 640
Db 1930 GGCATTTACTTCAAACTTGTCACAATGCAGACAAGAGGAATGAATTCAGTTAGAAAT 1989
Qy 641 AlaThrGlyLysSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSerGly 660
Db 1990 GCCACTGGTGAATCCAAAAGTGAAGTGGCTTTGGAATGCTCTCCAAAAGATTTCAGGG 2049
Qy 661 SerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyAsn 680
Db 2050 TCCAGTTTAATAAAGAGATCACTCGCAGGAGTATACATGTCACCAAGGCCAAGAC 2109

Qy 681 ArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrpArg 700
Db 2110 AGAAAGCTTGGTACAAAAGAGGACTTGAATGAGAATGTTCCCTTCAGTTTCCCTTCGAGG 2169
Qy 701 IleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAlaIle 720
Db 2170 ATTCTGAAGCTGAACCAACTGAATGCCTTATTTTGTGGTATATTTGTGCTATTT 2229
Qy 721 IleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePhe 740
Db 2230 ATAAACGGAGGCTGCAACACGACGATTTTCAATAATATTTTCAAGGATATAGGATCTTT 2289
Qy 741 ThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe 760
Db 2290 ACCGAGATGAGATCCTGAAACAAAACGACAGAATAGTAACATGTTTCTGTAITGTTT 2349
Qy 761 LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys 780
Db 2350 CTAGTCTTGGAAATTAATTTCTTTTATTAATATTTTCTCCAGGGCTTCACATTTGGCAAA 2409
Qy 781 AlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGln 800
Db 2410 GCTGGGGAGATCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTGAGACAG 2469
Qy 801 AspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla 820
Db 2470 GATGTGAGCTGTTTATGACCTAAACACACCACTGGAGCATTTGACAACCAAGGCTGCC 2529
Qy 821 AsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn 840
Db 2530 AATGATGCGGCTCAAGTTAAAGGGCTATAGGTTCCAGCTTGGCTCATACCCAGAAAT 2589
Qy 841 IleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeu 860
Db 2590 ATAGCAAAATCTTGGACAGGCAATTAATATCTTAATCTATGTTGGCAATTAACACTT 2649
Qy 861 LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMetLeu 880
Db 2650 TTACTCTTAGCAATTTGTACCCATCAATGCAATAGCAGAGTTGTTGAAATGAAATGTTG 2709
Qy 881 SerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGlu 900
Db 2710 TCTGGACAAGCACTCAAGATAAGAAAGAGCTAGAAAGAGCTGGGAAGATTGCTACAGAA 2769
Qy 901 AlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyrMet 920
Db 2770 GCCATCGAAAACCTCCGAACCTGTTGTTTCTTGTACTCGGAGCAGAAAGTTTGAATACATG 2829
Qy 921 TyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGly 940
Db 2830 TATGCACAGAGTTTCCAAAGTACCATAACAGAACTCTTTGAGGAAGACACATCTTCGGG 2889
Qy 941 ValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPhe 960
Db 2890 GTCTCATTTTCTATCACCCAGCAATGATGATTTTTCCTATGCTGGCTGTTCCGGTTT 2949
Qy 961 GlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPheSer 980
Db 2950 GGTGCTACTTGGTGGCAAAATGAGTTCATGAACCTTCAGGATGTTCTTTGGTATTCTCA 3009
Qy 981 AlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyrAla 1000
Db 3010 GCTATTGCTTTGGTGCCATGGCAGTGGGAGGTCAGTTCATTGCTCTCTGACTATGACC 3069
Qy 1001 LysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIleAsp 1020
Db 3070 AAAGCAAAAGTATCAGCAGGCCACGTCATCATGATCATTTGAATAAAACCCCTCTGATTGAC 3129
Qy 1021 SerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGlu 1040
Db 3130 AGCTACAGCCCTCAGCCCTCAAGCCAAATACGTTGGAAGAAATGTGACATTTAATGAG 3189

QY	1041	ValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeuGlu	1060
Db	3190	GTCTGTTCACTATCCACTCGACAGACATCCCGTCTCCAGGGGCTGAGCCTCGAG	3249
QY	1061	ValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrVal	1080
Db	3250	GTGAAAGAGGGCAGACGCTGCCCTCTGCTAGTAGCAGTGGCTGTGGGAAGACACAGTT	3309
QY	1081	ValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGlyLys	1100
Db	3310	GTTCACTCCTAGAGCGCTTCATGACCCCTTGGCTGCTCAGTCTCATTTGATGGCAAA	3369
QY	1101	GluIleLysHisLeuAsnValGlnTyrPheArgAlaHisLeuGlyLysValSerGlnGlu	1120
Db	3370	GAGATAAGACCTGAATGTCCAGTGGCTCCGAGCACCTGGGCATCTGCTCAGGAG	3429
QY	1121	ProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgVal	1140
Db	3430	CCCATCCTGTTGACTGCAGCATTCGCGAGCAATTCCTATGGAGACAACAGCGGGTC	3489
QY	1141	ValSerHisGluGluLeuMetGlnAlaAlaLysGluAlaAsnIleHisPheIleGlu	1160
Db	3490	GTATCATCAAGAGATTATGCAGGAGCGCAAGAGGCGCAACATACACCATTCATCGAG	3549
QY	1161	ThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGly	1180
Db	3550	ACACTCCTCGAGAAATACAAACACAGATAGGAGACAAGAACCCAGCTCTCTGGTGGC	3609
QY	1181	GlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeu	1200
Db	3610	CAGAAACAGCGCATTCGCATAGTCGCGCTCTGTTGTAGACAGCCTCATATTTGCTTTTG	3669
QY	1201	AspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeuAsp	1220
Db	3670	GATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTGTCCAAAGGCCCTGGAC	3729
QY	1221	LysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsn	1240
Db	3730	AAAGCCAGAGAGGCCGACCTGCATGTGATCGCCACCGCTGTGCCACCATCCAGAAT	3789
QY	1241	AlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGlnGln	1260
Db	3790	GCAGATTTAATAGTGGTGTTCAGATGGCAAGTCAAGAGCATGGCACATCAACAG	3849
QY	1261	LeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLysArg	1280
Db	3850	CTGTGCTCAGAAAGCATCTATTTTCCATGATCATGATGATGATGATGATGATGATGAT	3909
RESULT 3	AF269224	4045 bp DNA linear	SYN 04-JUL-2000
LOCUS	AF269224	Synthetic construct Canis familiaris his-tagged-multidrug	
DEFINITION	resistance glycoprotein gene, complete cds.		
ACCESSION	AF269224		
VERSION	AF269224.1	GI:8926216	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE	1 (bases 1 to 4045)		
AUTHORS	Roulet,A.C., Puel,O., Gesta,S., Alvinerie,M. and Pineau,T.J.		
TITLE	Analysis of dog MDRI p-glycoprotein		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 4045)		
AUTHORS	Roulet,A.C., Puel,O., Gesta,S., Alvinerie,M. and Pineau,T.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (19-MAY-2000) Pharmacology, INRA, 180 Chemin de		
FEATURES	Tournefeuille, BP3, Toulouse 31931, France		
source	Location/Qualifiers		
	1. 4045		
	/organism="synthetic construct"		
	/db_xref="taxon:32630"		
	/note="Original mRNA isolated from Beagle dog colon"		

CDS	101..3964	/protein_id="AAF81747.1"	
	/codon_start=1	/transl_table=11	
	/product="his-tagged-multidrug resistance glycoprotein		
	MDRI"		
	/db_xref="GI:8926217"		
	/translation="MHHHHDDPEGGRKGAENKFWKMGKKKKKKKKKKKKKPTVSTFA		
	MFYSNWLDRLYMLVGTMAAIHGAALPLMLVFGNMTDSFANAGISRNKTPFVINE		
	STNTQHTFNHLEEMTTVAYYSGIGAGVLVAAYIQVSWFCAAGROILKIRKQFF		
	HAIMROEIGHFDVHDYGEUNTKDDVSKINEGIDKIGMFTQSIATFTTGFIVGTR		
	GWLTVILAIASPVGLISRAIKAILSSSTDRELLAYAKAGAAVEVLAAIRTVIAFG		
	GKKELRYNKNLEEARIGIKKAITANISIGAAFLIIYASALAFWYGTSLVSEY		
	SICQVLTVFSLVIGAFSGQASPSFEAFANARGAAEYFKIIDNPKSDISYKSGHK		
	PDNIKGLEFNKVFSPSRKEVKILKGLNLKVSGOTVALVNSCGKSTVQLMQR		
	LYDPTDGMVCDIGDQIRTNVRLREITGVVSOEPLVFATTAENIRYRENTWDEI		
	EXAVKANAYDFIMKLPKDFTLVGERGQALSGGOKORAIARALVRNPKILLDEAT		
	SALDTSEAVQVALDKAKGPTTIVIAHRLSTVRNADYTAGDDGVIIVKEGNHDELM		
	KEKGIYFKLVTMOTRGNEIELENATGESKESDALESPKDSGSLIKRSTPRSLHA		
	POQDRKLGTKEDLNENPVSEFWIRILKLNSTEWPFYVVGIFCAILINGLQPAFSIF		
	SRITGIFTRDEDPETKRONSMFTLVILGIIISFTFFLQGFTHGAKGELILTKRUI		
	MVFISMLRQDVSWFDDPKNTGALTTRLANDAAQVKAIGSLAVITQNIANGTGI		
	ISLYGWQLTLLLAIVPIIAIAGVVMKMLSGOALKDKKEGAKIATEATENFT		
	VVSLTREQFEYMAQSLQVYRNSLRKAHIFGVFSITQAMMYFVAGCFRFGALIV		
	ANEFMFDQVLLVFSALVFCMAVAGQVSSFPADYAKAKYSAAHVIMLEKSPILDSY		
	PHGLKPNTELEGNVTNEVFNTPRPDIPVLQGLSLEVKGOTLALVSGGCKSTV		
	QLLERFYDLAGSVLIDGKEIKHLNVQWLRAHLGIVSQEPILFDCSTAEINTGNSR		
	VSHSEIMOAKKANIIHFTETPEKYNTRVGDKTQLSGGOKORAIARALVRPHI		
	LLDDEATSLDTESEKVVQEALDKAREGRTCIIVIAHRLSTIQNALSTVFPQNGKVEH		
	GTHQQLLAQKGIYFSLVQAGAK"		
BASE COUNT	1171 a 850 c 986 g 1038 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	0	Length:	4045
Score:	6428.50	Matches:	1274
Percent Similarity:	99.77%	Conservative:	3
Best Local Similarity:	99.53%	Mismatches:	2
Query Match:	99.40%	Indels:	1
DB:	12	Gaps:	1
US-09-672-725C-4 (1-1280) x AF269224 (1-4045)			
QY	2	AspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTyrLysMetGlyLysLys	21
Db	122	GATCTCTGAAGAGCGCGTAAGGGGAGTGCAGAGAAGAACTCTGGAAAATGGGCAAAAA	181
QY	22	SerLysLys---GluLysLysGluLysLysProThrValSerThrPheAlaMetPheArg	40
Db	182	AGTAAAAAATGAGAGAAAGAAAGAAACCACTGTCCAGCACGTTTGCATGTTTCCG	241
QY	41	TyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleHis	60
Db	242	TATTCAAATTTGGCTTGATAGGTGTATATGTTGGTGGGACCAATGGCTGCCATCATCCAT	301
QY	61	GlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAsn	80
Db	302	GGAGCTGCACCTCCCTCTCATGATGCTGGTTTTGGAAACATCACAGATAGCTTTCAAAAT	361
QY	81	AlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsn	100
Db	362	GCAGGAATTTCAAGAAACAAAACTTTTCCAGTTATAATTAATGAAGTATTACGAACAAAT	421
QY	101	ThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyrSer	120
Db	422	ACACAACATTTTCATCAACCATCTGGAGGAGGAATGACCAACATATGCTATATTTACAGT	481
QY	121	GlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeuAla	140
Db	482	GGGATCGGTGCTGGCGTCTGGTGGCTGTACATCCAGGTTTCATCTCTGTGTCCTGGCA	541
QY	141	AlaGlyArgGlnIleLeuLysLysArgLysGlnPhePheHisAlaIleMetArgGlnGlu	160

Db 542 GCAGGAAGACAGACTACTCAAAATTTAGAAAACAATTTTTTCATGCTATCATCGCACAGGAG 601
QY 161 IleGlyTyrPheAspValHisAspValGlyLeuAsnThrArgLeuThrAspAspVal 180
Db 602 ATTGGCTGTTTGACGTGCATCAGCTTGGGAGCTTAACACCCGGCTTCAGACGAGATGC 661
QY 181 SerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAlaThr 200
Db 662 TCCAAAATCAATGAAGAAATGGCGACAAAATTTGGAATGTTCTTCAATCAATAGCAACA 721
QY 201 PhePheThrGlyPheIleValGlyPheThrProGlyTyrLysLeuValIleLeu 220
Db 722 TTTTTCACCGGTTTTATAGTGGGGTTTACACGTGGTGGGAAAGTAAACCCCTGTGTGATTTG 781
QY 221 AlaIleSerProValLeuGlyLeuSerAlaAlaIleTyrAlaLysIleLeuSerSerPhe 240
Db 782 GCCATCAGCCCTGTTCTGGACITTCAGCCGCCATCTGGCAAGATATATCTTCATTT 841
QY 241 ThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValaGluGluValLeuAla 260
Db 842 ACTGATAAAGAACTTTGGCCTATGCAAAAGCTGGAGCAGTAGCTCAAGAAGTCTTAGCA 901
QY 261 AlaIleArgThrValIleAlaPheGlyGlnLysLysLysLysLysLysLysLys 280
Db 902 GCAATCAGAACTGTGATGGCTTTGGAGGACAAAAGAAAGAACTTGAAGGTACAAACAAA 961
QY 281 AsnLeuGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSerIle 300
Db 962 AATTTAGAAAGCTAAAGAAATTTGGATAAAGAAAGCTATCACGGCCCAACATTTCTATT 1021
QY 301 GlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrGlyThrSer 320
Db 1022 GGTGCCCTTCTTATGTGATCTATGCATCATATGCTCTGGCTTCTGGTATGGGACCTCC 1081
QY 321 LeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerValLeu 340
Db 1082 TTGGTCTCTCCAGTGAATATCTATGGCAAGTACTCAGCTGCTCTCTTTCTGTATTA 1141
QY 341 IleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArg 360
Db 1142 ATTTGGGGCTTTTATGATTTGGACAGGCATCCCAAGCAATTTGAAGCAATTTGCCAAACGCAAGA 1201
QY 361 GlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSer 380
Db 1202 GGAGCAGCTTATGAATCTTCAAGATAATTTGACAATAAACCAAGCAATTTGCAGAGCTATTTCG 1261
QY 381 LysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPhe 400
Db 1262 AAGAGTGGACATAAACCAGATAATTAAGGGAATTTGGAAATTCAAAATGTTCACTTC 1321
QY 401 SerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSer 420
Db 1322 AGTTACCTCTCGAAAAGAAAGTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTTCAGAGT 1381
QY 421 GlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeu 440
Db 1382 GGGCAGACAGTGGCGCTGGTGGGAACAGTGGCTCGGGAAGAGCACCGCTGCACCTG 1441
QY 441 MetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArg 460
Db 1442 ATGCAGAGGCTCTATGACCCACAGATGGCATGGTCTGCTATTGATGGACAGGACATTTAGG 1501
QY 461 ThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeu 480
Db 1502 ACCATAAATGTAAAGCATCTTTGGGAAATTTACTGTGTGTGTGAGTCAGGAGCCCTGTGTTG 1561
QY 481 PheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAspGlu 500
Db 1562 TTTGCCACCAGTAGCTGAAGAACATTCGCTATGGCCCGCGAAATGTCCACCATGGATGAG 1621
QY 501 IleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsnLys 520
Db 1622 ATTTAGAAAGCTGTTAAGGAAGCCCAATGCCTATCATTTTATCATGAACACTACCTAATAAA 1681

QY 521 PheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArgIle 540
Db 1682 TTTGACACTCTGGTTGGAGAGAGAGGGCCAGCTGAGTGGTGGACAGAAACAGAGAATC 1741
QY 541 AlaIleAlaArgAlaLeuValArgAspProLysIleLeuLeuLeuAspGluAlaThrSer 560
Db 1742 GCCATTGCTCGGGCCCTGGTTCGCAACCCCAAGATTTCTCTGCTGGATGAGCAACGTCA 1801
QY 561 AlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLysGly 580
Db 1802 GCTCTGGACACTGAAAGTGAAGCAGTGTTCAGTGGCCCTGGATGAAGCCAGAAAGGC 1861
QY 581 ArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla 600
Db 1862 CGGACTACCATTTGATAGTCTCATCGTTTGTCTACAGTTTCGTAAATGCCGATGTCATTGCT 1921
QY 601 GlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGluLys 620
Db 1922 GGTTCATGATGGAGTCAATTTGGAGAAGAGAAATCATGATCAACTCATGAAGAGAGAG 1981
QY 621 GlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGluAsn 640
Db 1982 GGCATTTACTTCAAACTTGTCACAATGCAGACAAGAGGAAATGAAATTCAGTTAGAAAAT 2041
QY 641 AlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSerGly 660
Db 2042 GCCACTGGTGAATCCAAAAGTGAAGTGAATGCTTGGAAATGCTCCAAAAGATTCAGGG 2101
QY 661 SerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAsp 680
Db 2102 TCCAGITTTAATAAAGAGATCACTCCGAGGATATACATGCACCAAGGCCCAAGAC 2161
QY 681 ArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTyrArg 700
Db 2162 AGAAAGCTTGGTACAAAGAGGACTTGAATGAGAATGAGATGACCTCCAGTTTCTCTTCGGAGG 2221
QY 701 IleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAlaIle 720
Db 2222 ATTCTGAGCTGACNCTCAACTGAATGGCCCTTATTTGTGGTGGTATATTTTGTGCTATT 2281
QY 721 IleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePhe 740
Db 2282 ATAAACGGAGGCTGCAACACGATTTTCAATAATATTTCAAGGATTTATAGGATCTTT 2341
QY 741 ThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe 760
Db 2342 ACCCGAGATGAGGATGCTGAAACAAACACAGAAATAGTAACATGTTTCTGTATTGTTT 2401
QY 761 LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys 780
Db 2402 CTAGTCTTGGAAATATTTCTTTTATACATTTTCTCCAGGGCTTCACATTTGGCAAA 2461
QY 781 AlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGln 800
Db 2462 GCTGGGAGATCTCTCACTAAGCGCTTCGATACATGTTTTCAGATCCATCCTCAGACAG 2521
QY 801 AspValSerTyrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla 820
Db 2522 GATGTCAGCTGGTTGTAGACCTTAAAGGCCATAGTTCCAGGCTTGTGTCATTTACCAGAA 2581
QY 821 AsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn 840
Db 2582 AATGATCGGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTGTGTCATTTACCAGAA 2641
QY 841 IleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeu 860
Db 2642 ATAGAAAATCTTGGACAGGCAATTTATATCTTAATCTATGTTGGCAATTAACACTT 2701
QY 861 LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMetLeu 880
Db 2702 TTACTCTTAGCAATTTGACCACCATTCATTCGAATAGCAGAGTGTGTTGAATGAAATGTTG 2761


```
Db 2237 TTTTACCCGAGATGAGGATCCTGAAACAAAACGACAGATAGTAACATGTTTCTGTATTG 2296
Qy 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnClyPheThrPheGly 779
Db 2297 TTTCTAGTCCTTGGAAATATTTCTTTTATATACATTTTCTCCAGGGCTTCACATTTGGC 2356
Qy 780 LysAlaGlyGluIleLeuThrLysArgLeuArgGlyTyrMetValPheArgSerMetLeuArg 799
Db 2357 AAGCTGGGGAGATCCTCACTAAGCGCTTCGATACATGGTTTTCAGATCCATGCTGAGA 2416
Qy 800 GlnAspValSerTrpPheAspPaspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
Db 2417 CAGGATGTCAGCTGGTTTGTATGACCCCTAAACACACACCTGGAGCATTTGACAAACCGGTT 2476
Qy 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
Db 2477 GCCAATGATGCGCTCAAGTTAAGGGCTATAGGTTCCAGGCTGCTGCTCAATACCCAG 2536
Qy 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 859
Db 2537 AATATAGCAAACTCTGGGACAGGCATTTATATATCCTTAATCTATGTTGGCAATTAACA 2596
Qy 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
Db 2597 CTTTTACTCTTAGCAATTTGACCATTTGCAATAGCAGGAGTTGTTGAATGAAATG 2656
Qy 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
Db 2657 TTGCTGGACACACACTGAAGATAGAAAGAGCTAGAGGAGCTGGGAAGATTGCTACA 2716
Qy 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
Db 2717 GAAGCCATCGAAACTTCGCACTGTTGTTCTTCTGACTCGGAGCAGAGTTTGAATAC 2776
Qy 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
Db 2777 ATGTATGCACAGAGTTTGCAGTACCATACAGAAATCTTTTGAGGAAGCACACATCTTC 2836
Qy 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
Db 2837 GGGGTCTCAATTTCTATCACCCAGCAATGATGTATTTTCCATGCTGGCTGTTCCGG 2896
Qy 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 979
Db 2897 TTTGGTGCCTACTTGGTGGCAATAGTTCATGAACCTTCAGGATGTTCTTTGGTATT 2956
Qy 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
Db 2957 TCAGCTATTGCTTTGGTGCCTAGCGCAGTGGGCGAGGTCAGTTCATTGCTCCTGACTAT 3016
Qy 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
Db 3017 GCCAAAGCCAAAGTATCAGCAGCCAGCTCATCATGATCATATGAAAAAGCCCTCTGATT 3076
Qy 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
Db 3077 GACAGCTACAGCCCTCAGCGCCTCAAGCCAAATACGTTGGAAGAAATGACATTTAAT 3136
Qy 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
Db 3137 GAGTGTGTGTTCACTATCCACTCGACACAGACATCCCGCTGCCAGGGGCTGAGCCCT 3196
Qy 1060 GluValLysLysGlyGlnThrLeuAlaLeuValIcylSerSerGlyCysGlyLysSerThr 1079
Db 3197 GAGGTGAAGAGGGCCAGACGCTGGCCCTCGTAGTAGCAGTGGCTGGCTGAGTGAAGACACA 3256
Qy 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
Db 3257 GTTGTTCAGCTCCCTAGAGCGCTTCTATGACCCCTTGGCTGGTTCAGTGCCTAATGATGGC 3316
Qy 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
Db 3317 AAGAGATAAAGCACTGAATGTCCAGTGGCTCCGAGCACACCTGGGCGATCGTCTCAG 3376

Qy 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
Db 3377 GAGCCCATCTCTGTTGACTGCAGCATTCGCCAGAACATTGCCATATGAGAGCAACAGCCGG 3436
Qy 1140 ValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPheIle 1159
Db 3437 GTCGTATACATGAGAGATTATGAGGAGCCCAAGGAGGCCAACATACACCATCTTCATC 3496
Qy 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
Db 3497 GAGACATCTCCCTGAGAATACACACAGAGTAGAGACAAAGAACCCAGCTCTCTGGT 3556
Qy 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
Db 3557 GGCCAGAAACAGCGCATTTGCCATAGCTCGCGCTCTTGTAGACAGCCTCATATTTGCTT 3616
Qy 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
Db 3617 TTGGATGAAGCTAGCATCTGATACAGAAAGTGAAGAGTTGTCGAAGAAGCCCTG 3676
Qy 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 3677 GACAAAGCCAGAGAGCGCCACCTGCATTGTCGCCACCGCTTGTCCACCATCCAG 3736
Qy 1240 AsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
Db 3737 AATGCAGATTTAATAGTGGTGTTCAGAAATGCAAGTCAAGAGGATGGCCACACATCAA 3796
Qy 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 3797 CAGCTGCTGGCCAGAGAGGATCTATTTTTCCATGGTGCAGTGTCCAGGCTGGAGCAAG 3856
Qy 1280 Arg 1280
Db 3857 CGC 3859

RESULT 5
AX105080 AX105080 4279 bp DNA linear PAT 30-APR-2001
LOCUS Sequence 24 from Patent WO0123540.
DEFINITION AX105080
ACCESSION AX105080
VERSION AX105080.1 GI:13921230
KEYWORDS
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 4279)
AUTHORS Stocker,P.J., Steimel-Crespi,D.T., Crespi,C.L., Reif,T.C. and
Patten,C.J.
TITLE P-glycoproteins and uses thereof
JOURNAL Patent: WO 0123540-A 24 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
source Location/Qualifiers
1. 4279
/db_xref="taxon:9615"
/organism="Canis familiaris"
/note="unnamed protein product"
17. 3862
/codon_start=1
/db_xref="GI:13921231"
/translaton="MDPEGPKGSAEKNFWKMKKSKKKEKKKPTVSTFAMPRSN
WLDRLYLVLGVNMAALIHGAALPLMLVLFGNMTDSFANAGISRNNKTFVILINESITNNT
QHFINHLEEMTYAYTYSIGAGVLAAYIQVSWFCLAAAGRLKLRKQFFHAIMQ
EIGWFDVHDVGLNTRLTDVSKINEGIDKIGMFFQSIATFFTFIVGTRGWLKL
VILASIPVLGSAALWAKILSSFTDKELLAYAKAGAAVEELAAIRTVAFGGQKKEL
ERYNKNEAKGIGIKKATANI SIGAFTLLIYASALAFWGTSLVLSSEYSGQVL
TVFVSILIGASIGQASPSIEAFANARCAVEIFKIIDNKP SIDYSKSGHKPNIKG
NLERKNVHFSIPSRKREVKILKGLNKLKVSQGTVALVGNCGGSKSTVQLMQLYDPTD
GMVCIDGDIETINVRHLREITGVVSEPLFATIIAETIRGENTMDEIEKAYKE
ANAYDFIMKLPKFTDITLGERGAQLSGGQKORIAIALVRNPKILLILDATSLDTE
```


SEAVVQVLDKARKGRRTTIVIAHRLSTVRNADYIAGFDDGVIVEKGNHDELMEKEGIY
FKLVTQTRNGNEIELENATGESKSEDALEMPKDSGLIKRRSTRRSITHAPQGDQR
KLGTKWEDNENPVSVFWRILKLNSTEDNPFVYVFGIFCAINGLQPAFSTIFSRIGI
FTRODEPETRQNSMFSVLVLGIIISFTPELOFTFKAGEIILTKRLRYVVFESM
LRQDSWFDDPKNTTCALTRLANDAAQVKGAIGSRNAVITONIANLGNIIISLYG
WQTLTLLALLVPJIAIAGVVMKMLSGOALKRKELEGAGKATIEATENFTVVSUTR
EQKFEMYAQLQVPYRNSLRKAHIFGVSFSTQAMMYTSYAGCFRFGAYLVANEFMN
FQDLVLFVSAIVFGAMAGQVVSFADYAKAKVSAAHVIMIIEKSPIDISYSPHGLKP
NTLENGVTFNEVVNPTRPDIPLVLOGLSLEVKKGOTLALVSGSGCGKSTVVOLLERF
YDPLAGSVLIDGREIKELHNQWLRHLGIVSORPILFDCSIAENIAYGDSNRVSHEE
IMQAAKEANIHFIETPLPEKYNTRVGDGTQLSGGQKQRIATARALVROPHILLDEA
TSALDTESEKVOQEAALDKAREGRTCTIVIAHRLSTIONADLVIVFVQNGKVEHGTHTOOL
LAQKGIYFMSVVOAGAKR*

BASE COUNT 1296 a 833 c 1008 g 1142 t
ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4279
Score: 6425.50 Matches: 1273
Percent Similarity: 99.69% Conservative: 4
Best Local Similarity: 99.38% Mismatches: 3
Query Match: 99.36% Indels: 1
DB: 6 Gaps: 1

US-09-672-725C-4 (1-1280) x AXI05080 (1-4279)

QY	1	MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys	20
DB	17	ATGGATCTCGAAGGAGCGCGTAAGGGAGTGCAGAGAAGAACTTCTGGAAATGGGCANA	76
QY	21	LysSer---LysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe	39
DB	77	AAAGTAAAAAAGAGAGAAGAAAGAAACCACTGTCAGCACGTTTGCNAATGTTT	136
QY	40	ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle	59
DB	137	CGCTATTCAAAATTCATCAACCATCTGTATATGTTGGTGGGACAAATGGCTGCATCATC	196
QY	60	HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	79
DB	197	CATGGAGTCGACCTCCCTCATGATGCTGTTTGGAAACATGACAGATAGCTTTGCA	256
QY	80	AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	99
DB	257	AATCGAGAAATTCAGAAACAAAACCTTTCAGTTATATTAATGAAGATTAACGAAC	316
QY	100	AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr	119
DB	317	AATACACAACATTTTCATCAACCATCTGGAGGAGAAATGACCAGTATGCTATTATTAC	376
QY	120	SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu	139
DB	377	AGTGGGATCGGTGCTGGCTGCTGCTTACATCCAGGTTTCATTCTGGTGCTG	436
QY	140	AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln	159
DB	437	GCACGAGAGACAGATACTCAAAATTAGAAAAATTTTTCATGCTATCATCGACAG	496
QY	160	GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp	179
DB	497	GAGATTGGCTGGTTGACGTGACGTGAGGTTGGGAGCTTAACACCCGCTCACAGAGAT	556
QY	180	ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla	199
DB	557	GTCCTCAAAATCAATGAAGAAATGGCGACAAAATGGAATGTTCTTCAATCAATAGCA	616
QY	200	ThrPheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle	219
DB	617	ACATTTTTCACCGGTTTATAGTGGGTTTACACCTGGTTGGAAGCTAACCCCTTGATTT	676
QY	220	LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaTyrIleLeuSerSer	239
DB	677	TTGGCCATCAGCCCTGTTCTTGGACTTTTCAGCCGCCATCTGGGCAAGATACTATCTTCA	736

QY	240	PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu	259
DB	737	TTTACTGATAAAGAACTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA	796
QY	260	AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn	279
DB	797	GCAGCAATCAGAACTGTGATTGGCTTTGGAGGACAAAAGAAAGAACTTCAAAAGGTACAAC	856
QY	280	LysAsnLeuGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer	299
DB	857	AAAAATTTAGAAAGCTAAGGAATTTGGGATAAAGAAAGCTATCACGCCCAACATTTCT	916
QY	300	IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr	319
DB	917	ATTGGTGGCGCTTCTTATTGATCTATGCATCATATGCTCTGGCTTCTTGGTATGGGACC	976
QY	320	SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal	339
DB	977	TCCTTGGTCTCTCCAGTGAATATTCTATTGGACAAGTACTCACTGCTCTTCTTTCTGTA	1036
QY	340	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	359
DB	1037	TTAAATTTGGGCTTTTATGATTGGACAGGATCCCCAAGCATTTCAAGCATTTGCCAACGCA	1096
QY	360	ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	379
DB	1097	AGAGGAGCAGCTTATGAAATCTTCAAGATAATGACAAATAAACCAAGCAATTCACAGCTAT	1156
QY	380	SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	399
DB	1157	TCGAAGAGTGGACATAAACACAGATAATTAAGGGAAATTTGGAATTTCAAAATGTTTAC	1216
QY	400	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	419
DB	1217	TTCAAGTATACCTTCTCGAAAGAAAGTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTTCA	1276
QY	420	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln	439
DB	1277	AGTGGCAGACAGTGGCGCTGGTGGGAACAGTGGCTGCGGGAAGACAGACCGTGCAG	1336
QY	440	LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	459
DB	1337	CTCATGCAGAGGCTCTATGACCCACAGATGGCATGGTCTGTATTGATGCAGACAGACATT	1396
QY	460	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	479
DB	1397	AGACCATAAATGTAAGGCATCTTCGGGAAATTAAGTGGTGGTGGTGGTGGTGGTGGTGG	1456
QY	480	LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	499
DB	1457	TTGTTTGGCCACCACGATAGCTGNAACATTCGCTATGGCCGCGGAAATGTCACCATGGAT	1516
QY	500	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	519
DB	1517	GAGATTGAGAAAGCTGTTAAGGAAGCAATGCTATGATTTTATCATGAAACTACCTAAT	1576
QY	520	LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnArg	539
DB	1577	AAATTTGACACTCTGTTGGAGAGAGGGGCCACGCTGAGTGGTGGAGAGAAACAGAGA	1636
QY	540	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	559
DB	1637	ATCGCCATTGCTCGGGCCCTGGTTGCAACCCCAAGATTTCTTCTGCTGGATGAGGCAAG	1696
QY	560	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	579
DB	1697	TCAGCTCTGACACTGAAAGTGAAGCAGTGGTTTCAGGTGGCCCTGGATGAAGCCAGAAA	1756
QY	580	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	599
DB	1757	GGCCGACCTACCATTTGATAGCTCATGTTGTTGCTACAGTTTCGTAATGCCGATGTCATT	1816

QY 600 AlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
|||||
Db 1817 GCTGGTTTTGATGATGGAGTCATCTGGGAAAGGAAATCATGATGAACCATGAAGAG 1876

QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
|||||
Db 1877 AAGGGCATTTACTTCAAACTTGTCAATGCGACATGCAAGAGGAAATGAAATTGAGTTAGAA 1936

QY 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
|||||
Db 1937 AATGCCACTGGTGAAATCCAAAAGTGAAGTGAATGCTTGGAAATGCTCTCCAAAAGATTCA 1996

QY 660 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
|||||
Db 1997 GGGTCCAGTTTAATAAAAAAGAGATCAACTCGCAGAGGTATATACATGCACCAAGGCCAA 2056

QY 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrp 699
|||||
Db 2057 GACAGAAAGCTTGGTACAAAAGAGGACTTGAATGAGATGTACCTCCAGTTTCTTCCTGG 2116

QY 700 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719
|||||
Db 2117 AGGATTTCTGAAGCTGAACCTCAACTGAATGGCCCTATTATTGTGGTTGGTATATTGTGCT 2176

QY 720 IleIleAsnGlyCysLeuGlnProIlePheSerIleIlePheSerArgIleIleGlyIle 739
|||||
Db 2177 ATTATAAAGCGAGGCGCTGCAACAGCAGCATTTTCAATAATATTTTCAAGGATTATAGGATC 2236

QY 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
|||||
Db 2237 TTTACCCGAGATGAGGATCCTGAACAAACAGCAGATAGTAATGTTTCTGTATGTG 2296

QY 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
|||||
Db 2297 TTTTCTAGTCTTGGAAATTATTTCTTTATTACATTTTCTCCAGGCTTTCACATTTGGC 2356

QY 780 LysAlaGlyCylLeuLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
|||||
Db 2357 AAAGCTGGGGAGATCCTCACTAAGCGCTTCGATATCATGGTTTTCAGATCCATGCTGAGA 2416

QY 800 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
|||||
Db 2417 CAGGATCTCAGCTGGTTGATGACCTTAANAACACCATGGAGCATTTGACAAACAGGCTT 2476

QY 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
|||||
Db 2477 GCCAATGATCGGCTCAAGCTTAAGGGGCTATAGGTTCCAGGCTGCTGTCATTTACCCAG 2536

QY 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 859
|||||
Db 2537 AATATAGCAAACTTTGGGACAGGCAATTATATATCTTAACTATGTTGGCAATTAAACA 2596

QY 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
|||||
Db 2597 CTTTTHACTCTTAGCAATTTGCCATCATTTGCAATGACAGGAGTTGTTGAAATGAAATG 2656

QY 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
|||||
Db 2657 TTGCTCTGGACAAGCACTGAAGATGAAGAGAGCTAGAGGAGCTGGGAAGATTGCTACA 2716

QY 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
|||||
Db 2717 GAAGCCATCGAAACATTCGAACTGTGTTTCTTTGACTCGGGAGCAAGAGTTTGAATAC 2776

QY 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
|||||
Db 2777 ATGTATGCACAGTTTGGCAATACCATACACAAACTCTTTGAGGAAAGCACACATCTTC 2836

QY 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
|||||
Db 2837 GGGGTCTCATTTTCTATCACCCAGCAATGATGATATTTTCTCTATGCTGGCTTTCGGG 2896

QY 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 979

Db 2897 TTTGGTCCCTACTTGGTGGCAAAATGAGTTTCATGAACATTTTCAGGATGTTCTTTTGGTATTC 2956

QY 980 SerAlaIleValPheGlyAlaMetAlaValIcylGlnValSerSerPheAlaProAspTyr 999
|||||
Db 2957 TCAGCTATTGCTTTGGTGCCATGGCAGTGGGCGAGGTCAAGTTCAATTTGCTCCTGACTAT 3016

QY 1000 AlalysAlalysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
|||||
Db 3017 GCCAAAGCCAAAGTATCAGCAGCCACGTCATCATGATCATTTGAAAAAGCCCTCTGATT 3076

QY 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
|||||
Db 3077 GACAGCTACAGCCCTCAGGCCCTCAAGCCAAATACGTTGGAAGGAATGTGACATTTAAT 3136

QY 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
|||||
Db 3137 GAGTCTGTTCACCTATCCACTCGACGACACATCCCGTGTCCAGGGGCTGAGCCCTC 3196

QY 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
|||||
Db 3197 GAGGTGAAGAGGCCAGACGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGAGACACA 3256

QY 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
|||||
Db 3257 GTTCTTCAGCTCCTAGAGCGCTTCTATGACCCCTTGGCTGGTTCAGTGTAAATGTATGGC 3316

QY 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
|||||
Db 3317 AAAGAGATAAGCACCTTGAATGCCAGTGGCTCCGASCACACCTGGGCATCGTGTCTAG 3376

QY 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
|||||
Db 3377 GAGCCCATCTCTGTGACTGCGCATTTCCGAGAACATTTGCCATATGGAGACAACAGCCGG 3436

QY 1140 ValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPheIle 1159
|||||
Db 3437 GTCGTATCACATGAAGAGATTATGACGAGCCCAAGGAGGCCAACATACACCATCTCATC 3496

QY 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
|||||
Db 3497 GAGACATCTCCTGAGAAATACACACCAGAGTAGGAGACAAGAACCCAGCTCTCTGTT 3556

QY 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
|||||
Db 3557 GGCCAGAAACAGCGCATTTGCCATAGCTCGCGCTCTTGTAGACAGCCTCATATTTTGCCTT 3616

QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
|||||
Db 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAGCCCTG 3676

QY 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
|||||
Db 3677 GACAAAGCCAGAGAGCCGACCTGCATTTGTATGCCCCACCGCTTGTCCACCATCCAG 3736

QY 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
|||||
Db 3737 AATCGATTTTAATAGTGGTGTTCAGAAATGGCAAGTCAAGGAGCATGGCACACATCAA 3796

QY 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
|||||
Db 3797 CAGTGTCTGGCCCAAGAGGATCTATTTTCCATGTCAGTGTCCAGGCTGGAGCAAG 3856

QY 1280 Arg 1280
|||
Db 3857 CGC 3859

RESULT 6

AX105082

LOCUS

DEFINITION

ACCESSION

VERSION

4279 bp

DNA

linear

PAT

30-APR-2001

Db 1337 CTGATGCAGAGGCTCTATGACCCACAGATGGCATGGTCTGTATTGATGGACAGGACATT 1396
Qy 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
Db 1397 AGGACCATAAATGTAAGGCATCTCGGAAATTAATCTGGTGGTGGTGCAGGAGCCTGTG 1456
Qy 480 LeuPheAlaThrThrIleAlaGluAsnIleAtgTyrGlyArgGluAsnValThrMetAsp 499
Db 1457 TTGTTTGGCCACCACGATAGCTGAAACATATCGCTATAGCCCGGCAAAATGTCCACATGAT 1516
Qy 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
Db 1517 GAGATTGAGAAAGCTGTAAAGAACCAATGCTATGATTTTATCATGAAACATACCTAAT 1576
Qy 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
Db 1577 AAATTTGACACTCTGTTTGGAGAGAGGGGCCAGCTAGTGGTGGACAGAAACAGAGA 1636
Qy 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
Db 1637 ATCGCCATTTGCTCGGCCCTGTTGCGCAACCCCAAGATTCCTCTGCTGGATGAGCAAG 1696
Qy 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
Db 1697 TCAGCTCTGCACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCTGGATAGGCCAGAAA 1756
Qy 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
Db 1757 GGCCTGACTACCAATGTGATAGCTCATCGTGTGTCTACAGTTTGGTAAATGCCATGTCAT 1816
Qy 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
Db 1817 GCTGGTTTGATGATGGAGTCATTGTGGAGAAAGGAATCATGATGAACATCATGAAAGAG 1876
Qy 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
Db 1877 AAGGGCATTTACTTCAAACTGTGCACAAATGCAGACAGAGGAATGAAATGAGTTAGAA 1936
Qy 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
Db 1937 AATGCCACTGGTGAATCCAAAGTGAAGATGATGCTTGGAAATGTCTCCAAAGATTCA 1996
Qy 660 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
Db 1997 GGGTCCAGTTTATAAAAGAAGATCAACTCGCAGGAGTATACATGCACACCAAGGCCAA 2056
Qy 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTyr 699
Db 2057 GACAGAAGCTTGGTACAAAAGAGGACTTGAATGAGAATGTACTCCAGTTTCCTCTGG 2116
Qy 700 ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 719
Db 2117 AGGATTCGAGCTGAACCTCACTGAATGCCCTATTATTTGTGGTGTATATTTGTGCT 2176
Qy 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
Db 2177 ATTTATAACGAGGCGCTCGCAACCAGCATTTTCAATAATATTTTCAAGGATATAGGGATC 2236
Qy 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
Db 2237 TTTACCCGAGATGAGGATCTCTGAAACAAAACGACAGATAGTAACATGTTTCTCTGATTG 2296
Qy 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
Db 2297 TTTCTAGTCTTGGAAATATTCTTTTATTATACATTTTCTCCAGGGCTTTCACATTTGGC 2356
Qy 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
Db 2357 AAAGCTGGGGAGATTCCTCACTAAACGCGCTCGATACATGTTTTCAGATCCATCGCTGAGA 2416
Qy 800 GlnAspValSerTyrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
Db 2417 CAGGATGTACGCTGGTTGTATGACCTTAAACACACCACCTGGAGCATTTGACACCGGCTT 2476

Qy 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
Db 2477 GCCAATGATCGGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTGCTGTTCATTACCAG 2536
Qy 840 AsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTyrGlnLeuThr 859
Db 2537 AATATAGCAATCTTGGGACAGCATTTATTATATCTTAACTATGTTGGTGGCAATTACA 2596
Qy 860 LeuLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLysMet 879
Db 2597 CTTTACTCTTAGCAATGTGACCATCATTTGCAATACAGGAGTGTGTAATGAAATG 2656
Qy 880 LeuSerGlyGlnAlaLeuLysAspLysGlyLeuGluGlyAlaGlyLysIleAlaThr 899
Db 2657 TTGCTGGCAAGCACTGAAAGATAAGAAAGAGCTAGAGGAGCTGGGAAGATTGTGTACA 2716
Qy 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
Db 2717 GAAGCCATCGAAAACTTCCGAACTGTTGTTCTTTGACTCGGAGACAGATTGTAATC 2776
Qy 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
Db 2777 ATGTATGCACAGATTTGCAAGTACCATACAGAAACTCTTTGAGGAAAGCACATCTTC 2836
Qy 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
Db 2837 GGGGTCTCATTTCTATCACCCAGGCAATGATGATTTTCTTATGCTGGCTGTTCCGG 2896
Qy 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 979
Db 2897 TTTGGTGGCTACTTGGTGGCAAAAGAGTTCATGAACITTCAGGATGTTCTTTTGGTATTC 2956
Qy 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
Db 2957 TCAGCTATTGCTTTGGTGCCATGGCAGTGGGAGGTTCAGTTTCATTGCTTCCTGACTAT 3016
Qy 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
Db 3017 GCCAAGCCAAAGTATCAGACGCCCATCATCATGATCATTTGAAAGAACCCCTCTGATT 3076
Qy 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
Db 3077 GACACTACAGCCCTCAGCGCTCAAGCCAAATAGTTTGGAGGAATGTGACATTTAAT 3136
Qy 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlySerLeu 1059
Db 3137 GAGGTGCGTGTCAACTATCCACTCGACGACATFCCCGTCTCCAGGGGCTGAGCGCTC 3196
Qy 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
Db 3197 GAGGTGAAGAGGGCCAGACGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGACACA 3256
Qy 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
Db 3257 GTTGTTCAGTCTCTAGAGCGCTTCATGACCCCTTGGTGGTTTCAGTCTCAATTGATGGC 3316
Qy 1100 LysGluIleLysHisLeuAsnValGlnThrPheArgAlaHisLeuGlyIleValSerGln 1119
Db 3317 AAAGAGATAAAGACCTGAATGTCAGTGGCTCCGAGCACACCTGGGCTCGTGTCTCAG 3376
Qy 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
Db 3377 GAGCCATCTGTTGACTGACGATTCGCGAGAACATTTGCTATGGAGACAACAGCGG 3436
Qy 1140 ValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPheIle 1159
Db 3437 GTCGTATCATGAAGATTTGTGAGGCGAGCAAGAGGCGCAACATACACCTTCATC 3496
Qy 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
Db 3497 GAGACACTCCCTGGAGAAATACAAACACACAGATGAGAGACAAGAACCCAGCTCTCTGT 3556

Qy 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
Db 3557 GCCAGAAACAGCGCATGGCATAGCTCGCGCTCTGTTAGACAGCCCTCATATTTGCTT 3616
Qy 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
Db 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGGTTGTCCAAGAAGCCCTG 3676
Qy 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 3677 GACAAACCCAGAGAGCCGACCTGCTGATGTCGCCACCCGCTGTGCCACCATCCAG 3736
Qy 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
Db 3737 AATGCAGATTAAATAGTGGTGTTCAGAAATGCCAAAGTCAAGGAGCATGCCACATCAA 3796
Qy 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 3797 CAGCTGCTGGCCAGAAAGCATCTATTTTCCATGTCAGTGTCCAGGCTGGAGCAAG 3856
Qy 1280 Arg 1280
Db 3857 CGC 3859
RESULT 7
AX105057
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source
CDS
dog.
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 4279)
Stocker,P.J., Stelmel-Crespi,D.T., Crespi,C.L., Reif,T.C. and
Patten,C.J.
P-glycoproteins and uses thereof
Patent: WO 0123540-A 1 05-APR-2001;
GENTEST CORPORATION (US)
Location/Qualifiers
1..4279
/organism="Canis familiaris"
/db_xref="taxon:9615"
17..3862
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37718.1"
/translation="MDPEGGRKSGAEKNFMWGMKSKKNEKKKPTVSTFAMPFRYS
WLDRLYMLVGTMAAIIHGAAALPLMLVFGNMTDSFANAGISRNKTFEPIVINESLTNNT
OHFINHLEEMTVAYIYSGIGAGVILVAAYIOVSFWCLAGROILKIRKQFFHLMRQ
EIGWFDHVDGELNRLTDDVSKINEGIDKIGMFFHSIAFTFFGIVGFTRGWKLTL
BILNLSPIVGSAAIWAKLISSTDKELLAYAKAAVEVLAIRVIAFGGOKKEL
VRYNNLEEAQGIKKAITANISIGAFLLIYASALAFWYGTSLVLSSEYSIGQVL
TVFFSLVIGASIGQASIEAFANARGAAVEIFEKIDNKPISIDYSKSGKPDNIKG
NLEFNKNHFSYSPSRKEVKILKLNKQSGOTVALVNSGCGKSTTVQLMRQDYPTD
GMVCIDGQDRTINVRHLREITGVVSOEPLVFATTIENRYGRENVTMDIEKAVKE
ANAYDFIMKLNKEDTLVGRGALSGGQKQRIARIALVRNPKILLILDEATSLDTE
SEAVVQALDKARKGRTTIVIAHRLSTVRNADVIAGDFDGVIVEKGNHDELMKGIY
FKLVTMGTNRNEIELENATGESDALEMSPKDSGLIKRSTRRSIHAPOQDR
KLGTKEOLNENPVSFWIRILKLNSTEWPFVVGIFCAIINGGLOPAFSLIFSRIIGI
LRQDVSFDDPKNMTGALTTRFLANDAAQVKAIGSLRAVITONIANLGTGIIISLIY
WQLRILLALLIPIITAIAGVVMKMSQALKKELEGAKIATELENFTVVSLTR
EQKEIFMIAQSLQVPIYNSURKAHIFGVFSITQAMMYFIAGCFRFGAYILVANEFN
FQDVLVFSALFVGAMAVQSVSEFAPYAKAKVAHVIMIIEKSPIDISYSPHGLKP
NTLEGNTFNEVNYNTPRDPVLOGLSLEVKKGTQLALVSGSGCGKSTVQVLLERF
YDPLAGSVLDGKEIKHLNQLWLRHGLIVSQEPILFDCSTAEINIAVGDNVSRVSEE
IMOAKKANITHHETIPEKYNTRVGDGKQLSGGQKQRIARIALRVQPHILLIDEA
TSALDTESEKVQBALDKREGRTICIVIAHRLSTIQNADLIIVFQNGKVEHETHOOL
LAQKGIYFSMVSVOAGAKR"

BASE COUNT 1294 a 834 c 1008 g 1143 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6420.50 Matches: 1272
Percent Similarity: 99.61% Conservative: 4
Best Local Similarity: 99.30% Mismatches: 4
Query Match: 99.28% Indels: 1
DB: 6 Gaps: 1
US-09-672-725C-4 (1-1280) x AX105057 (1-4279)
Qy 1 MetAspProGluGluGlyClyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 17 ATGGATCTCTGAAGGAGCGCGTAAGGGAGTGCAGAGAAGAACTTCTGGAAAATGGCAA 76
Qy 21 LysSerLysLys---GluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
Db 77 AAAGTAAAAAAAATGAGAAAGAAAGAAACCAACTGTCAGCACGTTTGCATGTTT 136
Qy 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 59
Db 137 CGCTATTCAAAATGGCTTGATAGTGTGATATGTTGGTGGGACAATGGCTGCCATCATC 196
Qy 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
Db 197 CATGGAGCTGCACTCCCTCTCATGATGCTGTTTGTGAAAACATGACAGATAGCTTTGCA 256
Qy 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
Db 257 AATGCAGGAATTCAGAAACAACAACTTTCAGTTATAATAAGTAAAGTATTACGAAC 316
Qy 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 119
Db 317 AATACACAACATTTTCATCAACCATCTCGGAGGAGGAATGACCAGTATGCCTATTATTAC 376
Qy 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaIleGlnValSerPheTrpCysLeu 139
Db 377 AGTGGGATCGGTGCTGGCGTCTGCTGCTGCTTACATCCAGGTTTCATTCGTGGCTG 436
Qy 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
Db 437 GCAGCAGGAAGACAGATACTCAAAATTAGAAAACAATTTTTCATGATCATCATCGGACAG 496
Qy 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
Db 497 GAGATTGGCTGGTTTGACGTGCATGACGTTGGGGAGCTTAACACCCGCTCACAGACGAT 556
Qy 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
Db 557 GTCTCCAAAATCAATGAAGGAATTTGGCGACAAAATTTGGAATGTTCTTCTACTCAATAGCA 616
Qy 200 ThrPhePheThrGlyPheIleValIcIlyPheThrProGlyTrpLysLeuThrLeuValIle 219
Db 617 ACATTTTTCACCGGTTTATAGTGGGTTTACACGTGGTGGGAAGCTAACCCCTGTGATT 676
Qy 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
Db 677 TTGGCCATCAGCCCTGTTCTTGGACTTTCAGCCGCCACTCTGGGCAAGATACTATCTTCA 736
Qy 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
Db 737 TTTTACTATAAAGAACCTCTTGGCCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTCTTA 796
Qy 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
Db 797 GCACCAATCAGAACTGTGATTGCTTTTGGAGGACAAAAGAAAGAACTTTGAAAGGTACAAC 856
Qy 280 LysAsnLeuGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
Db 857 AAAAATTTAGAAGAAGCTAAAGGAATTTGGGATAAAGAAAGCTATCATCGGGCAACATTCT 916

QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrTyrGlyThr 319
 Db 917 ATTGGTGGCGCTTCTTATGATCATGATATGCTGGCTTCTGGTATGGGACC 976
 QY 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
 Db 977 TCTTGGTCTCTCCAGTGAATATCTATGGACAAGTACTACATGCTCTCTTTCTGTA 1036
 QY 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAla 359
 Db 1037 TTAATTTGGGCTTTAGTATTTGGACAGCATCCCAAGCATTTGAAGCATTTGCAACGCA 1096
 QY 360 ArgGlyAlaAlaTyrCluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
 Db 1097 AGAGGAGCAGCTTATGAATCTTCAAGATTAATGCAATTAACCAAGCATTTGACAGCTAT 1156
 QY 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
 Db 1157 TCGAAGAGTGGACATAAACACAGATAATTAAGGGAAATTTGGAATTCAAAAATGTTTAC 1216
 QY 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
 Db 1217 TTTCAATTTACCTTCTCGAAAAGAGTAAAGATCTTAAAGGGTCTCAACCTGAAGGTTCAG 1276
 QY 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
 Db 1277 AGTGGCAGACAGTGGCTGGTGGGAACAGTGGCTGGGGAAGACAGCAGCCGTGCAG 1336
 QY 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
 Db 1337 CTGATGCAGAGGCTCTATGACCCACAGATGGCATGGTCTGTATTGATGGACAGGACAT 1396
 QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
 Db 1397 AGACCAATAAATGTAAGGCATCTTCGGGAATTTACTTGGTGGTGGTGGTGGTGGTGGT 1456
 QY 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
 Db 1457 TTGTTTGGCCACCATAGTATGAAACATTCGCTATGGCCGCGGAAATGTCACCATGGAT 1516
 QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
 Db 1517 GAGATTGAGAAAGCTGTATAGGAAGCAATGCTATGATTTTATCATGAACATACCTAAT 1576
 QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
 Db 1577 AAATTCACACTCTGGTGGAGAGAGAGGGGCCAGCTGAGTGGTGCACAGAAACAGAGA 1636
 QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
 Db 1637 ATCGCCATTGCTGGGCGCTTGGTTCGCAACCCCAAGATTTCTGCTGGATGAGGCAACG 1696
 QY 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
 Db 1697 TCAGCTTGACACTGAAAGTGAAGCAGTGGTTCAGTGGCCCTGGATTAAGGCCACAGAAA 1756
 QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
 Db 1757 GGCCGGACTACCATGTGATAGCTCATCGTTTGTCTACAGTTCTGTAATGCCGATGTCAAT 1816
 QY 600 AlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
 Db 1817 GCTGGTTTGTATGATGGAGTCATTTGGAGAAAGGAATCATGATGAATCAATGAAGAG 1876
 QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
 Db 1877 AAGGCATTTACTTCAAACTTGTTCACAAATGCACAGACAGAGAAATGAAATGAGTTAGAA 1936
 QY 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
 Db 1937 AATGCCACTGTGTAATCCAAAGTGAAGTATGCCTTGGAAATGTCTCCAAAGATTCGA 1996
 QY 660 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679

Db 1997 GGGTCCAGTTTAATAAAGAGATCAACTTCGAGGAGTATACATGCACACAGGCCAA 2056
 QY 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTyr 699
 Db 2057 GACAGAAAGCTTGGTCAAAAGAGGACTTGAATGAGAAATGACTCCAGTTTCTTCTGG 2116
 QY 700 ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 719
 Db 2117 AGGATTCGAAGCTGAACCTCACTGAATGCTTATTTTCTGGTGGTATATTTTGTGCT 2176
 QY 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
 Db 2177 ATTAATAACGAGGCTTCGCAACCCAGCATTTTCAATAATTTTCAAGGATATAGGATC 2236
 QY 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
 Db 2237 TTTACCCGAGATGAGGATCTCTGAACAAACAGCAGAAATAGTAACATGTTTCTGTATTG 2296
 QY 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
 Db 2297 TTTCTAGTCTCTTGAATATTTCTTTTATTACATTTTCTCTCCAGGCTTTCACATTTGGC 2356
 QY 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
 Db 2357 AAAGCTGGGAGATCTCTACCTAAGCGCTTCGATACATGGTTTTCAGATCCATGCTCAGA 2416
 QY 800 GlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuLeuThrThrArgLeu 819
 Db 2417 CAGATGTCACCTGGTTGATGACCTTAAACAAACACCACTGGAGCATTTGACAACCGAGCT 2476
 QY 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
 Db 2477 GCCAATGATCGGCTCAAGTTTAAAGGGCTATAGTTTCCAGGCTTGTGCTATTACCAG 2536
 QY 840 AsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTyrGlnLeuThr 859
 Db 2537 AATATAGCAATCTTGGACAGCATTTATATATCTTAACTATCTATGTTGSCAATAACA 2596
 QY 860 LeuLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLysMet 879
 Db 2597 CTTTTFACCTTAGCAATTTGACCATCATTCGAGGAGTGTGTGAATAAGAAATG 2656
 QY 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
 Db 2657 TTGCTGACAGCACTGAAAGATGAAGAGAGCTAGAAGAGCTGGGAAGATTCCTACA 2716
 QY 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
 Db 2717 GAAGCCATCGAAACTTCGAACTTCTTCTTTGACTCGGAGCAGAGATTTGAATAC 2776
 QY 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
 Db 2777 ATGTATGACAGAGTTTCAAGTACCATACAGAACTCTTTGAGGAAAGCAGACATCTTC 2836
 QY 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
 Db 2837 GGGGCTCATTTTCTATCATCCAGCAATGATGATTTTCTCTATGCTGGCTGTTCCGG 2896
 QY 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 979
 Db 2897 TTTGGTGCCTACTTGGTGGCAATGAGTTTCATGAACCTTCAGGATGTTCTTTTGTATTC 2956
 QY 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
 Db 2957 TCAGCTATTGCTTGGTGGCCATGGCAGGGGAGGAGTCACTCATTTCTTGTCTCTGACTAT 3016
 QY 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
 Db 3017 GCCAAGCCAAAGTATCAGCCAGCAGCTCATCATGATCATTTGAAAAAGCCCTCTGAT 3076
 QY 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039

Db	3077	GACAGCTACAGCCCTCAGCGCCTCAAGCCAAATACGTTGGAAGAAATGTGACATTTAAAT	3136
Qy	1040	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1059
Db	3137	GAGTCGTGTTCAACTATCCACTCGACACAGACATCCCGTGCTCCAGGGGTGAGCCCTC	3196
Qy	1060	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1079
Db	3197	GAGTGAAGAGAGGCCAGACGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGAGCACA	3256
Qy	1080	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1099
Db	3257	GTTGTTGAGCTCTTAGAGCGCTTCTATGACCCCTTGCTGGTTCAGTCAATGATGGC	3316
Qy	1100	LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1119
Db	3317	AAAGAGATAAGCACCTGAATGTCAGTGGCTCCGAGCACACTGGGCATCGTGTCTCAG	3376
Qy	1120	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1139
Db	3377	GAGCCCATCTCTGTTGACTGCAGCATGCGGAGAACATTGCCATTGGAGACAAACAGCCG	3436
Qy	1140	ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle	1159
Db	3437	GTCGTATCATGATGAAGAGATTATGCAGGCAGCCCAAGGAGGCCAACATACACCACTTCATC	3496
Qy	1160	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1179
Db	3497	GAGACATCCTCTGAGAAATACACACAGATAGGAGACAAAGGAACCCAGCTCTGTGGT	3556
Qy	1180	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1199
Db	3557	GGCCAGAAACAGCGCATTCGCCATAGCTCGCGCTCTGTTAGACAGCCTCATATTTGCTT	3616
Qy	1200	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1219
Db	3617	TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGGTTGTCCAAAGAGCCCTG	3676
Qy	1220	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1239
Db	3677	GACAAAGCCAGAGAAGCGCGACCTGCTGATGTGATCGCCACCGCTGTCCACCATCCAG	3736
Qy	1240	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1259
Db	3737	AATGCAGATTTAATAGTGTGTGTTTCAGATGCAAGTCAAGAGAGCATGGCACACATCAA	3796
Qy	1260	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys	1279
Db	3797	CAGCTCTGGCCCGAGAAAGGCATCTATTTTCCATGGGTCAAGTCCAGGCTGGAGCAAG	3856
Qy	1280	Arg 1280	
Db	3857	CGC 3859	
RESULT	8		
LOCUS	CFA419568		
DEFINITION	Canis familiaris mRNA for abortive p-glycoprotein (p-gp gene).		
ACCESSION	AJ419568		
VERSION	AJ419568.1		
KEYWORDS	p-glycoprotein; p-gp gene.		
SOURCE	dog.		
ORGANISM	Canis familiaris		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
TITLE	1 (sites)		
JOURNAL	Characterization of a MDR1a p-gp deficient dog in relation to		
REFERENCE	ivermectin sensitivity		
AUTHORS	Thesis (2002) Department of Biological Sciences, Universite de		
	Toulouse (UPS III), Toulouse, France		
	2 (bases 1 to 3934)		
	Roulet,A.		

TITLE	Direct Submission		
JOURNAL	Submitted (07-NOV-2001) Roulet A., Pharmacology, Inra, 180 Chemin de Tournefeuille, 31931 Toulouse cedex, FRANCE		
FEATURES	Location/Qualifiers		
source	1..3934		
	/organism="Canis familiaris"		
	/db_xref="taxon:9615"		
gene	/tissue_type="liver"		
	/note="ivermectin sensitive collies"		
	7..282		
	/gene="p-gp"		
CDS	7..282		
	/gene="p-gp"		
	/note="abortive product"		
	/codon_start=1		
	/evidence=experimental		
	/product="p-glycoprotein"		
	/protein_id="CAD12806.1"		
	/db_xref="GI:17385399"		
	/translation="MDPEGRRGSAENFWMGKSKKNEKKKPKTVSTFAMFRYSN		
	WLDRLYMLVGTMAAIIHGAAALPLMLLVFGNMTALQMGEFQETKLFQ"		
BASE COUNT	1144 a 815 c 964 g 1011 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	0 Length: 3934		
Score:	6385.00 Matches: 1270		
Percent Similarity:	99.45% Conservative: 4		
Best Local Similarity:	99.14% Mismatches: 5		
Query Match:	98.73% Indels: 3		
DB:	4 Gaps: 2		
US-09-672-725C-4 (1-1280) x CFA419568 (1-3934)			
Qy	1	MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys	20
Db	7	ATGGATCTCTGAAGGAGCGGTAGGGGGAGTGCAGAGAAGAACTCTGGAAGATGGCCAA	66
Qy	21	LysSerLysLys --- GluLysLysGluLysLysProThrValSerThrPheAlaMetPhe	39
Db	67	AAAGTAAAAAATGAGAAGAAAGAAAGAAACCAACTGTCAGCACGTTTGCATGTTT	126
Qy	40	ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle	59
Db	127	CCCTATTCAAAATGGCTGTAGGTGTATATGTTGGTGGGACAATGGCTGCCATCATC	186
Qy	60	HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	79
Db	187	CATGGAGCTGCACCTCCCTCTCATGATGCTGTTTTTGGAAACAT --- GACAGCTTGCA	242
Qy	80	AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	99
Db	243	AATGCAGGAATTCAGAAACAAACACTTTTCCAGTTATAATTAATGAAGTATTACGAAC	302
Qy	100	AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr	119
Db	303	AATACACAACATTCATCAACCATCTGGAGGAGAAATGCCACGATGCTCATTATTATAC	362
Qy	120	SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu	139
Db	363	AGTGGATCGGTGCTGGCGTCTGGTGGCTGTACATCCAGGTTTCATCTCGTGGCTG	422
Qy	140	AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln	159
Db	423	GCAGCAGAGACAGACTACTCAAAATTAGAAAAACAATTTTTCATGCTATCATGCGCAG	482
Qy	160	GluIleGlyTyrPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp	179
Db	483	GAGATTGGCTGGTTGACGTGCATGCTGGGGAGCTTAACACCCGGCTCACACAGAT	542
Qy	180	ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPheGlnSerIleAla	199
Db	543	GTCCTCAAAATCAATGAAGGAATCGGCGCAAAATTTGAATGTTCTTTCAATCAATAGCA	602

QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTyrLysLeuThrLeuValIle 219
 Db 603 ACATTTTTCACCGGTTTTATAGTGGGTTTACACGTGGTGGAGCTAACCCCTGTGTGATT 662
 QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaIleTyrAlaLysIleLeuSerSer 239
 Db 663 TTGGCCATCAGCCCTGTCTTGACCTTTCAGCCGCATCTGGCCAAAGATATCTTCA 722
 QY 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
 Db 723 TTTTACTGATAAAGACTCTTGGCTTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 782
 QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
 Db 783 GCAGCAATCAGAACTGTGATGCCCTTGGAGGACAAAGAAAGAACTTGAAGGTACAAAC 842
 QY 280 LysAsnLeuGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
 Db 843 AAAAATTTAGAGAGCTAAAGAAATTTGGATAAAGAAAGCTATCACGGCCCAACATTTCT 902
 QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrTyrGlyThr 319
 Db 903 ATTGGTCCGCTTCTTATTTGATCATGATCATATGCTCTGGCTTCTGGTATGGGACC 962
 QY 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
 Db 963 TCCCTGGCTCTCCAGTGAATATCTATTTGGACAAGTACTCAGTGTCTCTTCTCTGTA 1022
 QY 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
 Db 1023 TTAATTTGGGCTTTTATGATTGGACAGCATCCCAAGCATTTGAAGCATTTGCAACGCA 1082
 QY 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
 Db 1083 AGAGGAGCAGCTTATGAAATCTTCAAGATAATTTGAATTAACCAAGCATTCAGACCTAT 1142
 QY 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
 Db 1143 TCGAAGAGTGGACATAAACACAGATAATTAAGGGAATTTGGAATTTCAAAAATGTTAC 1202
 QY 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
 Db 1203 TTTCAATTACCTTCTCGAAAGAAGTTAAGACCTTAAAGGCTCTCAACCTGAAGGTTACG 1262
 QY 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
 Db 1263 AGTGGGCAGACAGTGGCCCTGGTGGGAACAGTGGCTGGGGAAGACGACCGCTGCAG 1322
 QY 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
 Db 1323 CTGATGACAGAGGCTCTATGACCCACAGATGGCATGGTCTGTATTGATGGACAGGACATT 1382
 QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrClyValValSerGlnGluProVal 479
 Db 1383 AGGACCAATAATGTAAGGCATCTTCGGGAAATTTACGTGGTGGTGGAGTCAGGAGCCTGG 1442
 QY 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
 Db 1443 TTGTTTCCACACGATAGCTGAAACATTCGCTATGGCCCGCAAAATGTCACCATTGAT 1502
 QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
 Db 1503 GAGATTGAGAAAGCTGTTAAGGAAGCCAAATGCCCTATGATTTTATCATGAAACTACCTAAT 1562
 QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
 Db 1563 AAAATTTGACACTCTGGTGGAGAGAGAGGGGCCCGCCAGCTGAGTGGTGACAGAAACAGAGA 1622
 QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
 Db 1623 ATCGCCATTGCTCGGGCCCTGGTTCGCAACCCCAAGATTTCTTCTGCTGGATGAGGCAACG 1682

QY 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
 Db 1683 TCAGCTCTGCACACTGAAAGTGAAGCAGTGGTTTCAGTGGCCCTGGATAGGCCAGAAAA 1742
 QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
 Db 1743 GGCGGGACTACCATGTTGATAGCTCATCGTTGTCTACAGTTCTGTAATGCCGATGTCATT 1802
 QY 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
 Db 1803 CTGTGTTTTGATGATGAGTCATTTGGAGAAAGAAATCATGATGAACATCATGAAAGAG 1862
 QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
 Db 1863 AAGGCACTTTACTTCAAACTTGTCCAAATGCAGACAAAGAGAAATGAAATTTGATTAGAA 1922
 QY 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
 Db 1923 AATGCCACTGGTGAATCCAAAGTGAAGTGAATGCCCTTGGAAATGCTCCAAAAGATTCA 1982
 QY 660 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
 Db 1983 GGGTCCAGTTTAAATAAAAGAAAGATCAACTCGCAGGAGTATACATGCACCAAGGCCAA 2042
 QY 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTyr 699
 Db 2043 GACAGAAAGCTTGGTACAAAGAGGACTTGAATGAGAAATGTACCTCCAGTTTCTCTCTGG 2102
 QY 700 ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 719
 Db 2103 AGGATTCGAAAGCTGAACCTCAACTGAATGGCTTATTTTGTGGTTGTATATTTGTGCT 2162
 QY 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
 Db 2163 ATTATAAACGGAGGCTCAACACGACATTTTCAATAATATTTTCAAGGATTTATAGGATC 2222
 QY 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
 Db 2223 TTTACCCGAGATGAGGATCCTGAAACAAACACGACAGATAGTAAACATGTTTCTGTATTG 2282
 QY 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
 Db 2283 TTTCTAGTCTTGGAAATTTATTTCTTTTATTTTCTTCTCCAGGGCTTCCATTTGGC 2342
 QY 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
 Db 2343 AAAGCTGGGAGATCCTCACTAAGCGGCTTCGATACATGGTTTTCAGATCCATGCTGAGA 2402
 QY 800 GlnAspValSerTyrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
 Db 2403 CAGGATGTCAGCTGGTGTGATGACCTTAAACACCACTGGAGCATTTGACAACCGGCTT 2462
 QY 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
 Db 2463 GCCAATGATCGGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGTGTTTATTTACCAG 2522
 QY 840 AsnIleAlaAsnLeuGlyThrClyIleIleIleSerLeuIleTyrGlyTyrGlnLeuThr 859
 Db 2523 AATATAGCAAAATCTTGGGACAGGCATTTATATCTTAACTATGTTGGCAATTAAACA 2582
 QY 860 LeuLeuLeuAlaIleValProIleIleAlaIleGlyValIleGluMetLysMet 879
 Db 2583 CTTTACTCTTAGCAATTTGTACCCATCATTCATAGCAGGAGTTGTTGAAATGAAATG 2642
 QY 880 LeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
 Db 2643 TTGTCTGGCAAGACACTGAAAGATAAGAAAGAGCTAGAGGAGCTGGGAAGATTCCTACA 2702
 QY 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
 Db 2703 GAAGCATCGAAACACTTCGAAACTGTTGTTCTTTGACTCGGACGACAGAGTTTGAATAC 2762
 QY 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939

```
Db 2763 ATGTATGCACAGAGTTTCAAGTACCAACAGAACTCTTTGAGGAAGCACACATCTTC 2822
Qy 940 GlyValSerPheSerIleThrGluAlaMetMetIleThrPheSerTyrAlaGlyCysPheArg 959
Db 2823 GGGGTCCTCATTTCTATCACCAGCAATGATGATATTTTCTCTATGCTGGCTGTTCGG 2882
Qy 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 979
Db 2883 TTTGGTGCTACTTGGTGCAATGAATTCATGAACCTTCAGGATGTTCTTTTGGTATTC 2942
Qy 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
Db 2943 TCAGCTATGTTCTTTGGTGCCATGCCAGTGGGCGAGTCACTTCATTTCCTCTGACTAT 3002
Qy 1000 AlalysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
Db 3003 GCCAAGCCAAAGTATCAGAGCCCGCATCATGATCATTTGAAAAAGCCCTCTGATT 3062
Qy 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
Db 3063 GACAGCTACAGCCCTCAGGCCCTCAAGCCAAATACGTTGGAAGGAATGTGACATTTAAT 3122
Qy 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
Db 3123 GAGTCTGTGTCTCACTATCCACTCGACACAGATCCCGTCTGCCAGGGGCTGAGCCTC 3182
Qy 1060 GluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
Db 3183 GAGGTGAAGAGGCCAGACGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGACACA 3242
Qy 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
Db 3243 GTTGTTCAGCTCTAGAGCGCTTCTATGACCCCTTGGCTGGTTCAGTGTCTAATGTATGC 3302
Qy 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
Db 3303 AAGAGATAAAGCACCTTGAATGTCCAGTGGCTCCGAGCACACTGGGCGATCGTGTCTCAG 3362
Qy 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
Db 3363 GAGCCATCTCTGTTGACTGCACCATTCGCGAGAACATTCGCTATGAGACACACCCGG 3422
Qy 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1159
Db 3423 GTCGTATCATGAAGAGATTATGCAGGAGCCAGGAGGCCAACATACACCACTTCATC 3482
Qy 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
Db 3483 GAGACACTCCTTGAGAAATACACACAGAGTAGGAGACAAAGGAAACCCAGCTCTCTGGT 3542
Qy 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
Db 3543 GGCCAGAAACAGCGCATTCGCCATAGCTCGCTCTGTTGAGACGCCCTCATATTTGCTT 3602
Qy 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
Db 3603 TTGGATGAAGCTACATCACTCGATACAGAAAGTGAAGAGTTGTCCAAGAGCCCTG 3662
Qy 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 3663 GACAAGCCAGAGAGGCCGACCTGCATTTGATCGCCACCGCTGTGTCACCATTCAG 3722
Qy 1240 AsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
Db 3723 AATGCAGATTTAATAGTGTGTTTTCAGAAATGCAAGTCAAGGACCATGGCACACATCAA 3782
Qy 1260 GlnLeuLeuAlaGlnLysGlyIleThrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 3783 CAGCTGCTGGCTCAAGAAAGGCATCTATTTTCCATGATCAGTGTCCAGGCTGGAGCAAA 3842
Qy 1280 Arg 1280
|||
```

```
Db 3843 CGC 3845
RESULT 9
AX322787
LOCUS AX322787
DEFINITION Sequence 1 from Patent WO0192877.
ACCESSION AX322787
VERSION AX322787.1 GI:18093766
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (sites)
AUTHORS Sorrentino,B. and Schuetz,J.
TITLE Method of identifying and/or isolating stem cells
JOURNAL Patent: WO 0192877-A 1 06-DEC-2001;
ST. JUDE CHILDREN'S RESEARCH HOSPITAL (US)
FEATURES
source
1..3860
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1135 a 746 c 957 g 1022 t
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3860
Score: 5859.50 Matches: 1159
Percent Similarity: 95.08% Conservative: 59
Best Local Similarity: 90.48% Mismatches: 60
Query Match: 90.61% Indels: 3
DB: 6 Gaps: 3
US-09-672-725C-4 (1-1280) x AX322787 (1-3860)
Qy 1 MetAspProGluGlyGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1 ATGGATCTTGAAGGGACCGCAATGGAGGAGCAAGAGAAAGAACTTTTAAACTGAAC 60
Qy 20 LysLysSerLysLysGluLysLysLysLysProThrValSerThrPheAlaMetPhe 39
Db 61 AATAAAGTGAAGAGATAAGAAAGAAAGAAACCAACTGTCAGTGATATTTCAATGTTT 120
Qy 40 ArgTyrSerAsnTrpLeuAspArgLeuThrMetLeuValGlyThrMetAlaAlaIleIle 59
Db 121 CCTATTCAATGGCTTGACAAGTTGATATGGTGGGAACTTTGGCTGCCATCATC 180
Qy 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
Db 181 CATGGGCTGGACTTCCTCTCATGATGCTGCTGTTGGAGAAATGACAGATATCTTTGCA 240
Qy 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
Db 241 AATGCAGGA---AATTTAGAAGACTCTGATGTCAAAACATCACTAATAAGAGTGATATCAAT 297
Qy 100 AsnThrGlnHisPheIleAsnHisLysGluGluMetThrThrTyrAlaTyrTyrTyr 119
Db 298 GATACAGGGTCTTCATGAAT---CTGGAGGAACATGACCATATGCTTATATTATC 354
Qy 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaIleIleGlnValSerPheTrpCysLeu 139
Db 355 AGTGAATATGGTCTGGGGTCTGCTGCTGCTTACATTCAGGTTTCATTTTGGTGCTG 414
Qy 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
Db 415 GCACCTGGAAGACAAAATACACAAAATAGAAAACAGTTTTTTCATGCTATAATGCGACAG 474
Qy 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
Db 475 GAGATAGCTGGTTTGTATGTGCAGATGTTGGGGAGCTTAACACCCGACCTTACAGATGAT 534
Qy 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPheGlnSerIleAla 199
|||||
```


Db 535 GTCTCTAAGATTAAATGAAGGTATTGGTGCACAAAATTTGGAATGTTCTTTTCAGTCAATGGCA 594
Qy 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTyrLysLeuThrLeuValIle 219
Db 595 ACATTTTTCACGCGGGTTTATAGTAGGATTTACACGTGGTTGGAAGCTAACCCCTTGTGATT 654
Qy 220 LeuAlaIleSerProValLeuGlyLeuSerAlaIleTyrPAlaLysIleLeuSerSer 239
Db 655 TTGGCCATCAGTCCTGTTCTTGACTGTTCAGCTGCTGCTGGCGAAAGATATACTATTCTCA 714
Qy 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
Db 715 TTTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTTGGACAGTAGCTGCAAGAGGCTTGG 774
Qy 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
Db 775 GCAGCAATTAGAACTCTGATTGCTATTGGAGGACAAAAGAAAGAACTTGAAAGGTACAAAC 834
Qy 280 LysAsnLeuGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
Db 835 AAAAATTTAGAAAGAACTTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCCAATATTCT 894
Qy 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrGlyThr 319
Db 895 ATAGGTGCTCTTCTGCTGATCTATGCATCTTATGCTCTGGCCCTTCGGTATGGGACC 954
Qy 320 SerLeuValLeuSerSerGlyTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
Db 955 ACCTTGGTCTCTCAGGGGAATATTCTATTGGACAAAGTACTCACTGATTTCTTTCTGTGA 1014
Qy 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
Db 1015 TTAATTCGGGCTTTTAGTGTGGACAGGCATCTCCAAAGCAATGAAGCAATTTGCCAATGCA 1074
Qy 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
Db 1075 AGAGGACAGCTTATGAATCTTCAAGATAATTGATAATAAAGCAAGTATTGACAGCTAT 1134
Qy 380 SerLysSerGlyHisLysLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
Db 1135 TCGAAGAGTGGGCACAAACCAAGATTAATTAAAGGAAATTTGGAATTCAGAAATGTTCCAC 1194
Qy 400 PheSerTyrProSerArgLysGlyValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
Db 1195 TTCAGTTACCCATCTCGAAAGAAGTTAAGATCTTGAAGGCGCTGAACCTGAAGGTGCAG 1254
Qy 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
Db 1255 AGTGGCAGACGGTGGCCCTGGTTGGAAACAGTAGTGGCTGGGGAAGACACACAGTCCAG 1314
Qy 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
Db 1315 CTGATGACAGAGGCTATATGACCCACAGAGGGAGTGTGAGTGTGATGACAGGATATT 1374
Qy 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
Db 1375 AGGACCAATAATGTAAAGGTTTACGGGAAATCATTTGGTGTGGTGAAGTGAAGACCTGTA 1434
Qy 480 LeuPheAlaThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
Db 1435 TTGTTTGGCCACGACGATAGCTGAAACAACTTCGCTATGGCGGTGAAATGTCACCATGGAT 1494
Qy 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
Db 1495 GAGATTGAGAAGCTGTCCAGGAAGCCATGCCATATGACTTATCATGAAGTGCCTCAT 1554
Qy 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
Db 1555 RAAATTTGACACCTCGTTGGAGAGAGAGGGGCCAGTTGAGTGTGGCGCAGAAAGCAGAGG 1614
Qy 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAlaThr 559
Db 1615 ATCGCCATTTGCACGTGCGCTGTTGCGCAACCCCAAGATCTCTCTGCTGATGAGGCCACG 1674

Qy 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
Db 1675 TCAGCCCTTGGACACAGAAAGCGAAGCAGTGGTTCAGGTGCTCTGGATAAGCCAGAAA 1734
Qy 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
Db 1735 GGTGGACCACTTGTGATAGCTCATCGTTGTCTACAGTTCGTAATGCTGACGTCATC 1794
Qy 600 AlaGlyPheAspGlyValIleValGlyLysGlyAsnHisAspGluLeuMetLysGlu 619
Db 1795 GCTGTTTCGATGATGGAGTCATTGTGGAGAAAGAAATCATGATGAACATCATGAAACAG 1854
Qy 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
Db 1855 AAAGCATTTACTTCAAACTTGTCAATGCAGACAGCAGGAAATGAAGTTGAATTAGAA 1914
Qy 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
Db 1915 AATGCAGCTGATGAATCCAAAAGTGAATTTGATGCTTGGAAATGCTTCAATGATTCATCA 1974
Qy 660 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
Db 1975 AGATCCAGTCTAATAAGAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAAGCCAA 2034
Qy 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTyr 699
Db 2035 GACAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCTCTTGG 2094
Qy 700 ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 719
Db 2095 AGGATTATGAAGCTAAATTTAACTGAATGGCTTATTTTGTGTGTGTATTTGTGGCC 2154
Qy 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
Db 2155 ATTATAATATGAGGAGGCTGCACACCACTTGCATAATATTTTCAAAGATATATAGGGT 2214
Qy 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
Db 2215 TTTTACAAGAATTTGATGATCTGAAACAAAACGACAGAAATAGTAACTGTTTTCATATTG 2274
Qy 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
Db 2275 TTTTCAGCCCTTGGAAATTAATTTCTTTTATACATTTTTCCTTCAAGGTTTTCACATTGGC 2334
Qy 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
Db 2335 AAAGCTGGAGAGATCTCACCAAGCGGCTCCGATACATGTTTTCGATCCATGCTCAGA 2394
Qy 800 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
Db 2395 CAGGATGTGAGTGGTGTGATGACCTAAACACACCTGGAGACATGTACTACAGGCTC 2454
Qy 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
Db 2455 GCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCAGGCTTGCCTGTAATATCCAG 2514
Qy 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTyrGlnLeuThr 859
Db 2515 AATATAGCAATCTTGGGACAGGAATAATATATCTTCTTCACTGTTGGCACTAAACA 2574
Qy 860 LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
Db 2575 CTGTTACTCTTAGCAATGTACCCTCATTTGCAATAGCAGGAGTGTGAAATGAAATG 2634
Qy 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
Db 2635 TTGCTGGCAACCACTGAAAGATAAGAAAGAACTAGAAGGTGCTGGGAAGATCGCTACT 2694
Qy 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
Db 2695 GAAGCAATAGAAACTTCCGAACCGTGTGTTCTTCTTACTCAGGAGCAGGAAGTTTGACAT 2754

Db 378 AATGCCAGGA---AATTTAGAGATCTGATGTCAAACATCACTAATAGAGTGATATCAAT 434
Qy 100 AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyrTyr 119
Db 435 GATACAGGGTTCTTCATGAAT---CTGGAGGAGACATGACCCAGGTATGCCTATATTATAC 491
Qy 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
Db 492 AGTGGAAATGGTCTGGGTGCTGTCTGCTTACATTCAGTTTCATTTTGGGCGCTG 551
Qy 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
Db 552 GCAGCTGGGAAGACAAATACACAAATATAGAAACAGTTTTTTCATATGCTATATGCGACAG 611
Qy 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
Db 612 GAGATAGCTGGTTTCATGTGCACGATGTTGGGGAGCTTAACACCCGACTTACAGATGAT 671
Qy 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
Db 672 GTCTCCAAGATTAAATGAAGGAATTTGGTACAAAATTTGGAATGTTCTTCAGTCAATGGCA 731
Qy 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
Db 732 ACATTTTCTACTGGTTTATAGTAGGATTTACACGTGTTGGAAGCTAACCCCTTGTGATT 791
Qy 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
Db 792 TTGGCCATCAGTCTCTGTGACTGTCTGACTGTCTGGCAAGATATCTTCA 851
Qy 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
Db 852 TTTTACTATAAAGAACTCTTAGCGTATGCAAAAGCTGGACGCTAGCTGAGAGGCTCTTG 911
Qy 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
Db 912 GCACCAATTAGAACTGTGATTGCAATTTGGAGGACAAAAGAAAGAACTTGAAGGTACAAC 971
Qy 280 LysAsnLeuGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
Db 972 AAAAATTTAGAAAGCTTAAAGAAATTTGGGATAAAGAAAGCTATACAGCCATATTTCT 1031
Qy 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
Db 1032 AFAGGTGCTGTTCTCTGCTGATCATGCACTTATGCTCTGGCTCTCTGATGGGACC 1091
Qy 320 SerLeuValLeuSerSerGlyTyrThrIleGlyGlnValLeuThrValPheSerVal 339
Db 1092 ACCTTGGTCTCTCAGGGGAATATTTCTATTGGACAAGTACTCACTGTATTCTTCTGTA 1151
Qy 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
Db 1152 TTAATTTGGGGCTTTTAGTGTGGACAGCATCTCCAGCATTTGAAGCATTTGCAAAATGCA 1211
Qy 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
Db 1212 AGAGGACGAGCTTATGAATCTTCAAGATAATTGATAATAAGCAAGTATTGACAGCTAT 1271
Qy 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
Db 1272 TCGAAGAGTGGGCACAAACACAGATAAATTAAGGAAATTTGGAATTCAGAAATGTTCCAC 1331
Qy 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
Db 1332 TTCAGTTACCCATCTCGAAAGAAAGTTAAGATCTTGAAGGCTCTGAACCTGAAGGTGCAG 1391
Qy 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
Db 1392 ACTGGGCACAGCTGGCCCTGTTGGAAACAGTGGCTGTGGAGAGAGACACACAGTCCAG 451
Qy 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
Db 1452 CTGATGCAGAGCTCTATGACCCACACAGAGGGATGGTCACTGATGGACAGGATATT 1511

Qy 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
Db 1512 AGGACCAATAATGTAAAGTTTCACGGGAAATCATTTGGTGGTGTAGTCAGGAACCTGTA 1571
Qy 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
Db 1572 TTGTTTGGCCACCAGATAGCTGAAACATTCGCTATGCGCGTGAAATGTGCAATGGAT 1631
Qy 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
Db 1632 GAGATTGAGAAGCTGTCAAGGAAGCCCAATGCTTACATTTATCATGAAATGCCTCAT 1691
Qy 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
Db 1692 AAATTTGACACCTTGTGGAGAGAGAGGGCCAGTTGAGTGTGGGCGCAAGACAGAG 1751
Qy 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
Db 1752 ATCGCCATTGCACGTGCCCTGTTTCGTAACCCCAAGATPCTCTGCTGGATGAGGCCACG 1811
Qy 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
Db 1812 TCAGCCTTGGACACAGAAAGCGACGCTGTTTTCAGGTGGCTCTGGATAAGGCCAGAAAA 1871
Qy 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
Db 1872 GGTGCGACCACTTGTGATAGTTCATCGTTTGTCTACAGTTCGTAATGCTGACGTCATC 1931
Qy 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
Db 1932 GCTGTTTTCGATGATGGAGTCAATGTGGAGAAAGAAATCATGATGAATCATGAAAGAG 1991
Qy 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
Db 1992 AAAGCAATTTACTTCAAACTTGTCACAATGCAGACAGCAGGAAATGAAGTTGAATTAGA 2051
Qy 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
Db 2052 AATGCACTGTGATGCAATCCAAAAGTGAATGATGCTTGGAAATGCTCTCAATGATTCA 2111
Qy 660 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
Db 2112 AGATCCAGTCTAATAAGAAAAAGATCACTCTAGTAGTGTCCGTGGATCACAAAGCCAA 2171
Qy 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrp 699
Db 2172 GACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGGTTTCCCTTTGG 2231
Qy 700 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719
Db 2232 AGGATTATGAAGCTAATTTAACTGAATGCGCTTATTTTGTGTTGTTGTTGTTGTTGTC 2291
Qy 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
Db 2292 ATTATAAATGGAGGCTGCAACCCAGCATTTGCAATAATATTTTCAAGATATTAGGGTT 2351
Qy 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
Db 2352 TTTTACAAGAAATTCATGATCTGAAACAAACAGACAGAAATAGTAATGTTTTCACATTATG 2411
Qy 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
Db 2412 TTTCTAGCCCTTGGAAATTTCTTTTATTACATTTTCTTCTTCTTCTTCTTCTTCTTCTT 2471
Qy 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
Db 2472 AAAGCTGGAGAGATCTCCACCAAGCGCTCCGATACATGTTTTTCCGATCCATGCTCAGA 2531
Qy 800 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
Db 2532 CAGGATGTGAGTGTGTTTGTATGATGCCCTTAAACACCACTGGAGCATTTGACTACCGGCTC 2591

QY	820	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	839
Db	2592	GCCATGATGCTGCTCAAGTAAAGGGCTATAGTCCAGGCTGCTGTAATACCCAG	2651
QY	840	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrglyTrpGlnLeuThr	859
Db	2652	AATATAGCAATCTTGGGACAGGAATAATATATCTTCATCTATGTTGGCAACTAACA	2711
QY	860	LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet	879
Db	2712	CTGTACTCTTAGCAATGTACCATCATTCGAATAGCAGAGTGTGGAATGAAATG	2771
QY	880	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr	899
Db	2772	TTGTCTGGCAAGCACTGAAGATGAAGAACTAGAGGTCTGGGAAGATCGCTACT	2831
QY	900	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTy	919
Db	2832	GAAGCAATAGAAAATCTCCGAACCGTGTCTTCTTGACTCAGGAGCAGAAATTTGAACAT	2891
QY	920	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	939
Db	2892	ATGTATGCTCAGATTTGCAGGTACCATACAGAACTCTTTGAGGAAGCACACATCTTT	2951
QY	940	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	959
Db	2952	GGATTAACATTTCTTCCACCCAGGCAATGATGATTTTCCATGCTGATGTTCCGG	3011
QY	960	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe	979
Db	3012	TTTGGAGCCACTTGGTGGGACATAAATACTCATGAGCTTTGAGGATGTTCTGTAGTATT	3071
QY	980	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	999
Db	3072	TCAGCTGTGCTTGTGGTCCATGGCGTGGGCAAGTCACTGATTTGCTCTGACTAT	3131
QY	1000	AlaLysAlaLysValSerAlaAlaHisValIleMetIleLeuLysSerProLeuIle	1019
Db	3132	GCCAAAGCCAAAATATCAGCAGCCACATCATCATCATTTGAAAAACCCCTTTGAT	3191
QY	1020	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1039
Db	3192	GACAGTACAGCAGGAGCCCTAATGCCGAACACATTTGGAGGAATGTACATTTGGT	3251
QY	1040	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1059
Db	3252	GAAGTGTATTCAACTATCCACCCAGCCGACATCCAGTCTTCAGGAGTACGCCG	3311
QY	1060	GluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1079
Db	3312	GAGGTGAAGAGGCGCAGACGCTGCTCTGGTGGCAGCAGTGGCTGTGGGAAGACACA	3371
QY	1080	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1099
Db	3372	GTGGTCCACTCTGGAGCGGTCTACGACCCCTTGGCAGGGAAGTGTGCTTATGGC	3431
QY	1100	LysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSerGln	1119
Db	3432	AAAGAAATAAAGCACTGAATTTCACTGGCTCCGAGCACACCTGGGCATCTGTGCCAG	3491
QY	1120	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1139
Db	3492	GAGCCCATCTCTTTGACGACGACATTTGCTGAGAATCTGCATGGAGACAACAGCCGG	3551
QY	1140	ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisIlePhe	1159
Db	3552	GTGGTGTACAGAGAGATTTGAGGCGCAGCAAGAGGCGCAACATACATGCTTCTATC	3611
QY	1160	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1179
Db	3612	GAGTCACTGCCTAATAAATATAGCACTAAAGTAGGACACAAGGAACCTCAGCTCTCGT	3671
QY	1180	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1199
Db	3672	GGCCAGAAACAGGCATTCATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	3731
QY	1200	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1219
Db	3732	TTGGATGAAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGAGTGTCCCAAGAGCCCTG	3791
QY	1220	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1239
Db	3792	GACAAAGCCAGAGAAGCCGACCTGCATTTGATGCTCAGCGCTGCTCCACCATCCAG	3851
QY	1240	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1259
Db	3852	AATGCAGACTTAATAGTGTGTTTTCAGAAATGCAGAGTCAAGAGCATGCCAGCATCAG	3911
QY	1260	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys	1279
Db	3912	CAGCTGCTGGCACAGAAAGCATCTATTTTCAATGGTCAGTGTCCAGGCTGGAAACAAG	3971
QY	1280	Arg 1280	
Db	3972	CGC 3974	
RESULT	11		
LOCUS	AX108654		
DEFINITION	Sequence 1 from Patent WO0123565.		
ACCESSION	AX108654		
VERSION	AX108654.1 GI:13923886		
KEYWORDS	crab-eating macaque.		
SOURCE	Macaca fascicularis		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae; Macaca.		
REFERENCE	1 (Stocker, P.J., 4186)		
AUTHORS	p-glycoproteins from macaca fascicularis and uses thereof		
TITLE	Patent: WO 0123565-A 1 05-APR-2001;		
JOURNAL	GENEST CORPORATION (US)		
FEATURES	Location/Qualifiers		
source	1..4186		
CDS	/organism="Macaca fascicularis"		
	/db_xref="taxon:9541"		
	100..3942		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC37764.1"		
	/db_xref="GI:13923887"		
	/translation="MDLEGDRNGAENKFFKLNKSKDKKPKKPKTVSVFSPFRYSN		
	WLDKLYMVVGTAAIHHGAGLPLMLVFGDMTDTFANAGNLGDLGALLTNSNITDTV		
	PVMLEEDMTRYAYYSIGAGVLVAAYIQVSWFLAAGROIHKKQFFHAIHQREI		
	GWFDVHDVGLNTRLDVSKINEGDKIGMFQSMATFTGIRKQFTGWLITVI		
	LAISPLVGLSAAVMWAKILSSFTDKELLYAKAGAAVEVLAARTVTFAGQKLEL		
	YKNLEAKRIGIKKATANTISGAFLLYASVALAFWYGTTLVLSKEYSIGQVLT		
	FEFLIGAFSGVQASPIEAFANAGAAFEIKLIIDNKSIDSYSKSGKPDNIKML		
	EERNVHESPSRKEYKILKGLNKVQSGQTVLVGSGCGKSTTVQLMQRLYDPTGEM		
	VSVDQDITINTVRFLEIIGVVSQEPVLFTATTAEINIRYGRDVTMDETEKAVKAN		
	AYDFIMKLPQKFDTLVGERGAQLSGGQKORIAARALVRNPKILLDEATSAIDTSE		
	AVQVALDKARKGRTTIVIAHRLSTVRNADVIAGDDGVIVKRSNDHLMKKEGIYFK		
	LVTMOTAGNEIELENADESKSEIDTLEMSHDSSGLIRKTRSRVSQSGQDKL		
	STKEALDESPVPSFWRIMKLNITEMPYFVVGVCALINGGLQPAFAVFSKILGIFT		
	RNDAEKTRQNSNLFLLFLVLGVIFIFFLQGTFFGAGLITLRLRTMVFPSMLR		
	QDVSFDDPNTTGLTTRIANDAAQVKAIGSLAIITQNIANDLGTIISLYIGMR		
	LTLLLIIVPITIAAGVEMKLSGALKDKKELEGAKIATEINFTVSLTQEQ		
	KFEHMTDQSLQVPRNSRKAHIFTFTSQAMTYFVAGCFRFAVLVMSLME		
	DVLVFSAVVFGAMAVQSSPADYAKAKVSAHIIIEIKTPLIDSYSTGLKPN		
	LEGNVTNEVVENYPTRLDIPVLQGLSLEKKGQTLALVSGSCGKSTVQLLEPYD		
	PLAGKVLDDKEIKQNLNVOVLAHLGIVSQEPLFDCISENIAVDGNRSRVSQEIV		
	RAAKENIHAFIESLPNKYTRVGDGTQLSGQKORIAARALVRNPKILLDEATS		
	ALDTSEKVVQVEALDKAREGRTCTIVIAHRLSTIQNADLIVFQNGRKEHGHQQLA		
	QKGIYFSWVSQAKRQ"		
BASE COUNT	1226 a 799 c 1039 g 1122 t		

ORIGIN

Alignment Scores:

Pred. No.:	0
Score:	5855.50
Percent Similarity:	95.32%
Best Local Similarity:	90.40%
Query Match:	90.54%
DB:	6

ПС-09-672-725С-4 (1-1280) x AX108654 (1-4186)

Qy	1	MetaspProGluGlyGlyArgLysGlySerAlaGlu--LysAsnPhetTrpLysMetGly	19
Db	100	ATGGATCTTGAGGGGACCCCAATGAGGAGCAGAGAGAACTTTTAACTGAAC	159
Qy	20	LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe	39
Db	160	AATAAAGTAAAAAGATAGAGGAAAGGAAACCAACTGTCAGTGATTTTCAATGTTT	219
Qy	40	ArgTyrSerAsnTrpLeuAspLeuTyrMetLeuValGlyThrMetAlaAlaIleIle	59
Db	220	CGCTATTCAAATGGCTTGACAAAGTTGTATATGGTGGCGAACTTTGGCTGCCATCATC	279
Qy	60	HisGlyAlaAlaLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	79
Db	280	CATGGAGCTGCACCTCTCATGATGCTGGGTGTTGGAGACATGACGATACCTTTGCA	339
Qy	80	AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	99
Db	340	AATGCAGGA--AATTATGGAGATTTAGGAGCTCTGTTCACTAATAATGACGATAATATCACT	396
Qy	100	AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr	119
Db	397	GATACAGTGGCCGTCATGAAT---CTGGAGGAAGATATGACCGATGTCCTATTATTATAC	453
Qy	120	SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu	139
Db	454	AGTGGAAATGGTGGGTGCTGGTGTGCTGTACATTACAGTTTTCATTTTGGTGGCGTG	513
Qy	140	AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln	159
Db	514	GCAGCTGGAAGACAAATACACAAATTTAGAAAAACAGTTTTTTCATGCTATAATGGCAGAC	573
Qy	160	GluIleGlyTtpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp	179
Db	574	GAGATAGCTGGTTGATGTGCACGATGTTGGGAGCTTAACACCCGGCTTACAGATGAT	633
Qy	180	ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPheGlnSerIleAla	199
Db	634	GTCTCCAAGATTAATGAGGAATTTGGTGACAAATTTGGAATGTCTTTCAGTCAATGGCA	693
Qy	200	ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle	219
Db	694	ACATTTTTTCATCGGCTTTTATAGTAGATTTACAGCTGGTTGGAAAGCTAACCTTTGTGAT	753
Qy	220	LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer	239
Db	754	TTGGCCATCAGTCCTGTCTTGGACTGTGAGCTGCAGCTGGGCAAGAGACTGTCTTCA	813
Qy	240	PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu	259
Db	814	TTTACTGATAAGAACTCTTACGTATGCAAAAGCTGGAGCAGTAGCTGGAAGAGTCTTGG	873
Qy	260	AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn	279
Db	874	GCAGCAATTAGAACTGTGATTTGCATTTGGAGGACAAAGAAAGAACTCGAAAGGTCAAC	933
Qy	280	LysAsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer	299
Db	934	AAAAATTTAGAAGAAGCTAAAAGAAATTCGGGATAAGAAAGCACTATTACAGGCAATTTCT	993
Qy	300	IleGlyAlaAlaPheLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr	319

Db	994	ATAGGTGCTGCTTCTGCTTATCTATCATCTTATGCTCTGGCCTTCTGGTATGGGACC	1053
Qy	320	SerLeuValLeuSerSerGluThrThrIleGlyGlnValLeuThrValPhePheSerVal	339
Db	1054	ACCTTGGTCTCTCAAGAAGATATTCATTTGGACAAGTACTCACTGATTTCTTTTCTGTGA	1113
Qy	340	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	359
Db	1114	TTAATTGGGGCTTTAGTCTTGGACAGGCATCTCCAAGCATTTGAAGCATTTGGCAATGCA	1173
Qy	360	ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	379
Db	1174	AGAGGACGACGTTTGAATCTTCAAGATAATGTGATAATAACCCAGTATTGACAGCTAT	1233
Qy	380	SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	399
Db	1234	TCGAAGAGTGGGCACAAACAGATAATATTAGGGGAAATTTGGAAATTCAGAAATGTTCCAC	1293
Qy	400	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	419
Db	1294	TTCAAGTATACCCATCTCGAAAGAAGTTAAGATCTTGAAGGGCCCTGAACCTGAAGGTGCAG	1353
Qy	420	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln	439
Db	1354	AGTGGGCAGACGCTGGCCCTGGTTGGAAACACGCGCTCTGGGAAGAGCACAAACGTCACG	1413
Qy	440	LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	459
Db	1414	CTGATGACAGAGGCTTATGACCCACAGAGGCATGGTCAGTTGTATGGACAGAGATTT	1473
Qy	460	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	479
Db	1474	AGGACCATAAACGTAGAGTTCTTACCGGGAAATCATCGGCTGGTGAGTCAGAACTGTA	1533
Qy	480	LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	499
Db	1534	TTGTTTGGCCACCAGATAGCTGAAAAACATTCGCTATGTGCTGAAGATGTCCACATGGAT	1593
Qy	500	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	519
Db	1594	GAGATTGAGAAGCGTCTCAAGGAAGCAATGCTATGACTTTATCTGGAACCTGCCTCAG	1653
Qy	520	LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg	539
Db	1654	AAATTTGACACCTCGTTGGAGAGAGAGGGGCCACGCTCAGTGGTGGGCAGAAAGCAGAGG	1713
Qy	540	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	559
Db	1714	ATCGCAATTGCACGTGCGCTTGGTTCGCAACCCCAAGATCCTCTGCTGGACGAGGCCACG	1773
Qy	560	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	579
Db	1774	TCAGGCTTGGACACAGAAAGTGAAGCAGTGGTTTCAGTGGCTCTGGATGAAGCCAGAAAA	1833
Qy	580	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	599
Db	1834	GGTCGGACCCACCATTTGTATAGCTCATPCGTTTCTCTACGCGTTCTGTAATGCCAGCTCATC	1893
Qy	600	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	619
Db	1894	GCTGGTTTCGATGATGGAGTCAATTTGGAGAAAGAAATCATGATGAGCTCATGAAGAG	1953
Qy	620	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	639
Db	1954	AAAGGCATTTACTTCAAACTTGTCACAATGCACACAGCAGGAAATCAAAATTTGAATTAAGAA	2013
Qy	640	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	659
Db	2014	AATGCAGCTGATGAATCCAAAGTGAAATTCAGTACCTTGGAAATGCTTTCACATGATCA	2073
Qy	660	GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln	679

```
Db 2074 GGATCCAGCTCTAATAAGAAAAGATCCACTCGTAGGAGTGTCCGTGGATCACAGGCCAA 2133
QY AsparGlyLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrp 699
Db 2134 GACAAAGCTTAGTACCAAGAGGCTGTGGATGAAGATATACCTCCAGTTTCCTTTGG 2193
QY ArgLeuLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719
Db 2194 AGGATTATGAAGCTAAATTTAACTAGCTGGCCCTATTATTTGTTGGTGATTTTGTGCC 2253
QY IleIleAsnGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
Db 2254 ATTATAATGAGGCTCGCAACACGACATTTGCAGTAATATTTCAAGATTATAGGATT 2313
QY PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
Db 2314 TTTACAAGAAATGATGATCGCGAACAACAGCAGATAGTAACTTGTTTTCCTACTATTG 2373
QY PheLeuValLeuGlyIleIleSerPheIleThrPheLeuGlnGlyPheThrPheGly 779
Db 2374 TTTCTAGCTCTGGAATTTGTTCTTTTATTATACATTTTTCCTTCAGGCTTCACATTTGGC 2433
QY LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
Db 2434 AAGCTGGAGAGATCTTCACCAAGCGGCTCCGATACATGGTTTTCGATCCTCATCTCAGA 2493
QY GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
Db 2494 CAGGATGTAGCTGGTTGATGACCTTAAACACACCCTGGAGCATTGACTACAGGCTC 2553
QY AlaAsnAspAlaIleGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
Db 2554 GCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCATATATACCCAG 2613
QY AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 859
Db 2614 AATATAGCAAACTTGGGACAGAAATTAATATATCTTAATCTATGTTGGCAACTGACA 2673
QY LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
Db 2674 CTGTTACTCTAGCAATTTGACCATCATTCGAATAGCAGAGTGTGTTGAATGAAATG 2733
QY LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
Db 2734 TTGCTCTGGACAAGCACTGAAGATAGAAAGAACTAGAAAGTGTCTGGGAAGATCGTACT 2793
QY GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
Db 2794 GAAGCAATAGAAAACCTCCGAACTGTTCTTTCTTTGACTCAGGAGCAAGTTTGAACAT 2853
QY MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
Db 2854 ATGTATGATCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGCAAGCACACATCTTT 2913
QY GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
Db 2914 GGAATCACGTTTCTCTCAGCAGCAATGATGATTTTCTATGCTGGATGTTCCGG 2973
QY PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 979
Db 2974 TTTGGAGCCTACTGCTGGCACAATAGTCTCATGAGCTTTGAGGATGTTCTGTAGTATT 3033
QY SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
Db 3034 TCAGCTGTGTCTTTGGTGCCATGCGCTGGGCAAGTCAGTTCAATTTGCTCCTGACTAT 3093
QY AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
Db 3094 GCCAAGCCAAAGTATCAGACGCCACATCATGATCATTTGATTAAGAAACCCCTTGATT 3153
QY AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
Db 3154 GACAGCTACAGCAGAGGCTTAAGCCGAACACATTTGGAAGAAATGTACATTTAAT 3213
```

```
QY 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
Db 3214 GAAGTTGTATTCAACTATCCACCCGACTGGACATCCCAAGTCTTCAGGGCTGAGCCTG 3273
QY 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
Db 3274 GAAGTGAAGAAGGCGCAGACGCTGGCCCTGGTGGCGACAGTGGCTGTGGGAAGACAG 3333
QY 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
Db 3334 GTGGTCCAGCTCCTGGAGCGGTTCTATGACCCCTTGGCGGGGAAAGTCTGCTTACGGC 3393
QY 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyLleValSerGln 1119
Db 3394 AAAGAAATAAAGCAACTGAATGTTGAGTGGCTCCGAGCACACCTGGCATCTGTCGCCAG 3453
QY 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
Db 3454 GAGCCCATCTCTTTCAGTGCAGCATTTAGTGAACAATTTGCCCTATGGAGACAAACAGCCGG 3513
QY 1140 ValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPheIle 1159
Db 3514 GTGGTGTACAGGAAGAGATCGTGAGGCAGCCAAAGGAGGCCAATATACACCCCTTCATC 3573
QY 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
Db 3574 GAGTCACTGCTCTTAATAATATATAGACCAGTAGGAGACAAAGGAACCTCAGCTCTCTGGT 3633
QY 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
Db 3634 GCCAAGAAACACGCAATTTGCCATAGCTGTGCCCTTGTGTAGACAGCCCTCATATTTTGT 3693
QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGlnAlaLeu 1219
Db 3694 TTGGATGAAGCCACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAGAGACCCCTG 3753
QY 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 3754 GACAAAGCCAGAAAGGCGGTACCTGCATTTGATTGTCACCCGCTGTCCACCATCCAG 3813
QY 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
Db 3814 AATGCAGACTTAATAGTGTGTTTCAGAAATGCGAGAGTCAAGGAGACGCGCACACATCAG 3873
QY 1260 GlnLeuAlaGlnLysGlyIleTyrPheSerMetLysValGlnAlaGlyAlaLys 1279
Db 3874 CAGCTGCTGGCAGCAAGAGGCATCTATTTTCAATGCTCAGTGTCCAGGCTGGAGCAAG 3933
QY 1280 Arg 1280
Db 3934 CGC 3936
```

```
RESULT 12
AX108656 LOCUS
DEFINITION Sequence 3 from Patent WO0123565.
ACCESSION AX108656
VERSION AX108656.1 GI:13923888
SOURCE crab-eating macaque.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Macaca.
REFERENCE 1 (bases 1 to 4195)
AUTHORS Stocker,P.J., Steimel-Crespi,D.T. and Crespi,C.L.
TITLE P-glycoproteins from macaca fascicularis and uses thereof
JOURNAL Patent: WO 0123565-A 3 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
    source
        location/Qualifiers
            1..4195
                /organism="Macaca fascicularis"
```

```

/db_xref="taxon:9541"
100..3951
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37765.1"
/db_xref="gi:13923889"
/translation="MDLEGDRNGAEKKNFKNLNNKSKDKKERRKPTVSFVSMFRYSN
WDKLIWVVGTLAAIIGAGLPLMWLVFGDMTDFANAGLGLDGLALLPNTNNSNIT
DTPVMNLDEMTRIAYISGIGAGVLAAYIQVSWCLAAAGROIHKIRKQFPAIMR
QIEGFVDVHDGELNLTDDVSKINEGIDKIGMFFOSMATFFIGFVIGTGWKLIT
LVILAISPLGLSAAVAKILISFTDKELLAYAGAVEVLAARIVLAFGQCKKE
LERKNLEAKRIGIKAITANSIGAFTLLIYASALAFWGTITVLVSKEYSIGOV
LTVFNSLIFAGFSGOASIEAFANARGAFFIKIIDNKPISIDSKSHKPDNIK
GNLERNVHFSYPSKKEVKILKGLNLVQSGOTVALVNSCGKSTVQLMORLYDPT
EGMVSDGQDITKINRFLREIIGVWQEPVLFATTAEINRVRGREDVTWDELEKAVK
EANAYFTDKLPQKDFDLVBERGAQLSGGQKQRIATAIRNRPNKIILLBERALSALT
ESEAVALQDARKKGRFTTIVIAHRLSTVRNADVIAGFDGVIIVEKGNHDELMEKGI
YFKLVTMOAGNETELENAADESKIDTLEMSSHDSGSLIRKRSRVRSGQGD
RKLSTPEALDESIPVSWFRIMKLNLTWEPYFVGVFCALINGLQPAFAVFSKIIG
IFTRNDRAETKRONLSFLLVLVIGIVSFTFFLOGFTFGKAGEILTNRVYVFRS
MLQRVDSFDDPKNTGALTTRLANDAAQVKGAGSLAIITONIANLGTGLIISLYI
QWQLTLLIILAIPIIAGVEMKLSGOALKDKKELEGAGKIATEALENFRVVSILT
QEKFEHMYDSLOQVYRNSLRKAHIFGITEFTQAMMYFSIACFRFAYLVAHSLM
SPEDVLLVFSVAFAMAVGVSSFPADYAKARVSAAHIMIEIKTPLDISYSTEGKL
PNTLEGNVTNEVFNYPTRLDIPVLGLSLEVKKGQTLALVSGSGCKSTVQGLLR
PYDPLAGKVLIDGKEIKOLNQMWRHLGIVSQBPILFDCSISENIAYGDNRSVSOE
ETVRAAKKANITHATESIPNKYSTPVGDKGTQLSGGOKORAIARALVRPHILLDE
ATSLDTESEKVVQALDKAREGRTCVIAHRLSTIQNALDILVVFQNGRVKHEGTHQO
LUAQKGIYFSMVQAGAKRQ"
BASE COUNT      1230 a      801 c      1039 g      1125 t
ORIGIN

Alignment Scores:
Pred. No.:      0      Length:      4195
Score:          5853.00      Matches:      1160
Percent Similarity: 94.80%      Conservative: 61
Best Local Similarity: 90.06%      Mismatches: 53
Query Match:      90.51%      Indels: 14
DB:                6      Gaps: 4

US-09-672-725C-4 (1-1280) x AX108656 (1-4195)
QY 1 MetaspProGluGlyGlyArgLysGlySerAlaGlu---LysAsnPhetPrLysMetGly 19
Db 100 ATGGATCTTGAAGGGGACCCCAATGGAGGAGCAGAGAAAGAACTTTTAACTTGAAC 159
QY 20 LysLysSerLysLysGluLysGlyLysProThrValSerThrPheAlaMetPhe 39
Db 160 AATAAAGTAAAGAAAGATAAGAAAGAAAGAAACCAACTGCTCAGTGTATTTCAATGTT 219
QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 220 CGCTATTCAAAATGGCTTGACAAAGTTCTATATGCTGGTGGAACTTTGGTGCCATCATC 279
QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
Db 280 CATGGAGCTGGACTTCCTCATGATGCTGGTGTGAGACATGACGGATACCTTTGCA 339
QY 80 AsnAlaGly-----IleSerArgAsnLysThrPheProValIle 92
Db 340 AATCAGGAATAATTTAGGAGATTTAGGAGCTCTGTGTTTAAACAACACT----- 387
QY 93 IleAsnGluSerIleThrAsnAsnThrGlnHisPheIleAsnHisLeuGluGluMet 112
Db 388 ---AATAGCAATAATATCACTGATACAGTCCCGCTCATGAAT---CTPGAGGAGATGATG 441
QY 113 ThrThrTyrAlaTyrTyrSerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIle 132
Db 442 ACCAGGTATGCTTATATACAGTGGAAATGGTGCTGGGGTGCTGGTGTGCTTACATT 501
QY 133 GlnValSerPheTrpCysLeuAlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhe 152
Db 502 CAGGTTTCATTTTGGTGCTGGCAGCTGGAAGACAAATATACAAAAATTAGAAACAGTTT 561

```

```

153 PheHisAlaIleMetArgGlnGluIleGlyTrpPheAspValHisAspValGlyGluLeu 172
156 TTTTCATGCTATATGCGACAGGAGATAGCGTGTGTTGATGTGCACGATGTGGGGAGCTT 621
173 AsnThrArgLeuThrAspAspValSerLysIleAsnGluGlyIleGlyAspLysValGly 192
622 AACACCCGGCTTACAGATGATGCTCCAGATTAAATGAAGAAATTTGTTGACAAATTTGA 681
193 MetPhePheGlnSerIleAlaThrPhePheThrGlyPheIleValGlyPheThrProGly 212
682 ATGTTCTTTCAGTCAATGGCAACATTTTTCACCTGGGTTTATAGTAGGATTTACACGTGT 741
213 TrpLysLeuThrLeuValIleLeuAlaIleSerProValLeuGlyLeuSerAlaAlaIle 232
742 TGGAAAGTAAACCTTGTGATTGTTGGCCATCAGTCTCTTCTTGGACTGTGTCAGTGCAGTC 801
233 TrpAlaLysIleLeuSerSerPheThrAspLysGluLeuAlaTyrAlaLysAlaGly 252
802 TGGCAAGATGACTGTCTTCAATTTACTGTATAAGAACTCTTACCTATGCAAAAAGCTGA 861
253 AlaValAlaGluGluValLeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLys 272
862 GCAGTAGCTGAAGAGTCTTGGCAGCAATTAGAACTGTGATTGCTATTTGGAGGACAAAAG 921
273 LysGluLeuGluArgTyrAsnLysAsnLeuGluGluAlaLysArgIleGlyIleLysLys 292
922 AAAGAACTCGAAAGGTACAAACAAAATTTAGAGAACTTAAAGAAATTTGGATTAAGAAA 981
293 AlaIleThrAlaAsnIleSerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAla 312
982 GCTATTACAGCCAAATATTTCTATAGTGTCTTCTTCTGCTTATCTATGCTATCTATGCT 1041
313 LeuAlaPheTrpTyrGlyThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnVal 332
1042 CTGGCCCTTCTGCTATGGGACCACTTGGTCTCTCAAGAGAAATATCTTATTGGACAAAGTA 1101
333 LeuThrValPhePheSerValIleGlyAlaPheSerIleGlyGlnAlaSerProSer 352
1102 CTCACGTGATTCTTTCTGCTATTAATTTGGGCTTTTAGTGTGGACAGGCATCTCCAAGC 1161
353 IleGluAlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsn 372
1162 ATTTGAAGCATTTGCAATGCAAGAGAGAGCAGCTTTTGAATCTTCAAGATTAATTTGAATA 1221
373 LysProSerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsn 392
1222 AAGCCAAAGTATTGACAGCTATTTCGAAGAGTGGGCAACCAACAGATTAATTTAAGGAAAT 1281
393 LeuGluPheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeuLys 412
1282 TTGGAATTCAGAAATGTTCACTTCAGTTACCCATCTCGAAAAGAAAGTTAAGATCTTGAAG 1341
413 GlyLeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCys 432
1342 GGCCTGAACCTGAAGGTGCGAGAGTGGGCGCTGGTGGGAAACAGCGGCTGT 1401
433 GlyLysSerThrThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetVal 452
1402 GGGAGAGCACAACCGTCCAGCTGATGAGAGGCTTTATGACCCACACAGAGGGCATGTGTC 1461
453 CysIleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThrGly 472
1462 AGTGTGATGGACAGGATATTAGGACCATAAACGTAAGGTTTCTACGGGAAATCATCGGT 1521
473 ValValSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGly 492
1522 GTGGTGAGTCAGAAACCTGTATGTTTGGCCACCATAGCTGAAACACATTCGCTATGCT 1581
493 ArgGluAsnValThrMetAspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAsp 512
1582 CGTGAAGATGTCACCATGATGAGATTGAGAAAGCTGTCAAGAAAGCAATGCCATGAC 1641

```

QY 513 PheIleMetLysLeuProAsnLysPheAspThrLeuValGluArgGlyAlaArgLeu 532
Db 1642 TTTATCATGAAATGCCCTCAGAAATTTGACACCTGGTTGGAGAGAGGGGCCAGCTG 1701
QY 533 SerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIle 552
Db 1702 AGTGGTGGCGAGAGACAGAGATGCCATTGCACGTGCCCTGGTTCCCAACCCCAAGATC 1761
QY 553 LeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValGlnVal 572
Db 1762 CTCTGCTGGAGAGAGCCCGCTGAGCCTTGACACAGAAAGTGAAGCAGTGGTTTCAGGTG 1821
QY 573 AlaLeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThr 592
Db 1822 GCTCTGGATAAGGCCAGAAAGGTCGACACCACATTGTGATAGCTCATCGTTGTCTACG 1881
QY 593 ValArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGluLysGlyAsn 612
Db 1882 GTTCGTAATGCCGACGTCATCGCTGGTTTCGATGATGGAGTCATTGTGGAGAAAGGAAT 1941
QY 613 HisAspGluLeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArg 632
Db 1942 CATGATGAGTCATGAAGAAGAGAGGCATTACTTCAAACTTGTCAAAATGCGACAGCA 2001
QY 633 GlyAsnGluIleGluLeuGluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeu 652
Db 2002 GGAATGAAATTTGAATAGAAATGCAAGTCGATGATCCAAAGTGAATTTGATACCTTG 2061
QY 653 GluMetSerProLysAspSerGlySerSerLeuIleLysArgArgSerThrArgArgSer 672
Db 2062 GAATGTCCTTCATGATGATCCAGATCCAGTCAGTCTAATAGAAAGATCCACCTCGTAGAGT 2121
QY 673 IleHisAlaProGlnGlyLysAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsn 692
Db 2122 GTCCGTGGATCACAGGCCAAGACACAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGT 2181
QY 693 ValProSerValSerPheThrArgIleLeuLysLeuAsnSerThrGluTrpProTyrPhe 712
Db 2182 ATACCTCCAGATTTTCCCTTTGGAGGATTATAGCTAAATTAACCTCAGTGGCTTATTTT 2241
QY 713 ValValGlyIlePheCysAlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIle 732
Db 2242 GTTGTGGTGATTTTGTGCCATTATAAATGAGGTCCTGCAACCCAGCATTTGCAGTAATA 2301
QY 733 PheSerArgIleIleGlyIlePheThrArgAspGluAspProGluThrLysArgGlnAsn 752
Db 2302 TTTTCAAGATTTATAGGATTTTACAGAAATGATGATGCCGAAACAAACAGAGAT 2361
QY 753 SerAsnMetPheSerValLeuPheLeuValLeuGlyIleIleSerPheIleThrPhePhe 772
Db 2362 AGTAACCTGTTTTCATATTGTTTCTAGTCCTTGAATTTGTTCTTTTATTACATTTTTC 2421
QY 773 LeuGlnGlyPheThrPheGlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMet 792
Db 2422 CTTCAGGGCTTCACATTTGGCAAGCTGGAGAGATCCCTCACCAGCGGCTCCGATACATG 2481
QY 793 ValPheArgSerMetLeuArgGlnAspValSerTrpPheAspAspProLysAsnThrThr 812
Db 2482 GTTTTCCGATCCATGCTCAGACAGGATGTGAGCTGGTTTGTATGACCCCTAAAACACCACT 2541
QY 813 GlyAlaLeuThrThrArgLeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySer 832
Db 2542 GGAGCATTTACTACCGGCTGCCATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCC 2601
QY 833 ArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeu 852
Db 2602 AGGCTTGCTATAATTTACCAAGATATAGCAATCTTGGACAGAGGAATTAATATATCCCTTA 2661
QY 853 IleTyrGlyTrpGlnLeuThrLeuLeuLeuAlaIleValProIleIleAlaIleAla 872
Db 2662 ATCTATGTTGGCAACTGACACTGTTACTCTTAGCAATTTGACCAATCATTCATGCAATGCA 2721
QY 873 GlyValValGluMetLysMetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGlu 892

Db 2722 GGAGTTGTTGAAATGAAATGTTCTCTGGACAAAGCACTGAAAGATAAGAAAGCACTAGAA 2781
QY 893 GlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeuThr 912
Db 2782 GGTGCTGGGAGATCGCTACTGAAGCAATAGAAACTTCCGAACCTGTTGTTCTTTGACT 2841
QY 913 ArgGluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSer 932
Db 2842 CAGGAGCAGAAGTTTCAACATATGTATGATCAGAGTTTCAGGTACCATACAGAACTCT 2901
QY 933 LeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyrPhe 952
Db 2902 TTGAGAAAGCACACATCTTTGGAATACGTTTTTCCCTCACGCAAGCAATGATGATATTT 2961
QY 953 SerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPhe 972
Db 2962 TCCATATGCTGGATGTTCCGGTTTGGAGGCTACTTGGTGGCACATAGTCTCATGAGCTTT 3021
QY 973 GlnAspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGlnVal 992
Db 3022 GAGGATGTTCTGTTAGTATTTTTCAGCTGTTGCTTTGGTGCCATGGCCGTGGGCAAGTC 3081
QY 993 SerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMetIle 1012
Db 3082 AGTTCATTTGCTCCTGACTATCCAAAGCAAGTATCAGCAGCCACATCATCATGATC 3141
QY 1013 IleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThrLeu 1032
Db 3142 ATTTGAAAACCCCTTGTATGACAGCTACACAGAGGCTTAAGCCGAACACATTTG 3201
QY 1033 GluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIlePro 1052
Db 3202 GAAGGAAATGTACATTTAATGAAGTTGTTTCAACTATCCACCCTGACTGGACATCCCA 3261
QY 1053 ValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySer 1072
Db 3262 GTGCTTCAGGGCTCAGGCTTGAAGTGAAGAGGCCAGACGCTGGGCCCTGGTGGGACG 3321
QY 1073 SerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeuAla 1092
Db 3322 AGTGGCTCTGGGAAGACACGCTGTGTCCAGCTCTCTGGAGCGTTCCTATGACCCCTTGGCG 3381
QY 1093 GlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAla 1112
Db 3382 GGGAAAGTGCTGTGTCGCGCAAGAAATAAGCAACTGAATGTTCAGTGGCTCCGACGA 3441
QY 1113 HisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIle 1132
Db 3442 CACCTGGGCATCGTGTCCAGAGGCCCATCTCTGTTGACTGCACATATAGTGAGAACATT 3501
QY 1133 AlaTyrGlyAspAsnSerArgValValSerHisGluGluIleMetGlnAlaLysGlu 1152
Db 3502 GCCTATGGAGACAAACAGCCGGTGTGTACAGGAAGAGATCGTGAGGGCAGCCCAAGAG 3561
QY 1153 AlaAsnIleHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAsp 1172
Db 3562 GCCAATATACAGCCTTCATCGACTCCTGCTTAATAAATATAGCACACAGTAGGAGAC 3621
QY 1173 LysGlyThrGlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuVal 1192
Db 3622 AAAGGAATCAGCTCTCTGTGGCCAGAAACAGCAATGCCATAGTCTCGGCCCTGTT 3681
QY 1193 ArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGlu 1212
Db 3682 AGACAGCCTCATATTTTGTCTTTGGATGAAGCCACATCAGCTCTGGATACAGAAAGTAA 3741
QY 1213 LysValValGlnGluAlaLeuAspLysAlaArgGluArgThrCysIleValIleAla 1232
Db 3742 AAGTTGTCACAGAGCCCTGGACAAAGCCAGAGAGGCCGTACCTGCAATGTTGATGCT 3801
QY 1233 HisArgLeuSerThrIleGlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysVal 1252

Db 3802 CACCCTGTCACCATCCAGATCCAGACTTAATAGTGGTGTTCAGATGGCAGAGTC 3861

Qy 1253 LysGluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetile 1272

Db 3862 AAGGAGCAGCGCACACATCAGCAGCTGCTGGCACAGAAAGGATCTATTTTCAATGGTC 3921

Qy 1273 SerValGlnAlaGlyAlaLysArg 1280

Db 3922 AGTGTCAGGCTGGAGCAAGCGC 3945

RESULT 13

AX322789

LOCUS AX322789 3860 bp DNA linear PAT 07-JAN-2002

DEFINITION Sequence 3 from Patent WO0192877.

ACCESSION AX322789

VERSION AX322789.1 GI:18093767

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (sites)

TITLE Sorrentino, B. and Schuetz, J.

JOURNAL Method of identifying and/or isolating stem cells

ST. JUDE CHILDREN'S RESEARCH HOSPITAL (US)

FEATURES

source

1..3860

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 1135 a 746 c 956 g 1023 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3860

Score: 5850.50 Matches: 1158

Percent Similarity: 95.00% Conservative: 59

Best Local Similarity: 90.40% Mismatches: 61

Query Match: 90.47% Indels: 3

DB: 6 Gaps: 3

US-09-672-725C-4 (1-1280) x AX322789 (1-3860)

Qy 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19

Db 1 ATGGATCTTGAGGGGACCCCAATGGAGGAGCAAGAGAAAGAACTTTTAAACTGAAC 60

Qy 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39

Db 61 AATAAAAGTGAAGAAAGATAAGAGAAAGAAACCAACTGTCTAGTGTATTTTCAATGTTT 120

Qy 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 59

Db 121 CGCTATTCAATATGGCTTGACAGTGTATATGGTGGGAACTTTGGCTGCCATCATC 180

Qy 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79

Db 181 CATGGGCTGCATCTCTCATGATGCTGTGTGTTGGAGAAATGACAGATATCTTTGCA 240

Qy 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99

Db 241 AATGCAGGA---AATTTAGAGATCTGATGTCAACATCATCAATAGAGTGATATCAAT 297

Qy 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrThrAlaTyrTyrTyr 119

Db 298 GATACAGGGTCTTCATGAAT---CTGGAGGAGACATGACAGATATGCTATTATATAC 354

Qy 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139

Db 355 AGTGAATTTGGTGGGGTCTGCTGCTTACATTCAGGTTTCATTTTGTGGCTG 414

Qy 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159

Db 415 GCAGCTGGAAGACAAATACACAAAATTAGAAAACAGCTTTTTCATGCTATTAATGCGACAG 474

Qy 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179

Db 475 GAGATAGGCTGGTGTGATGTCAGGATGTTGGGAGCTTAAACACCGACTTACAGATGAT 534

Qy 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199

Db 535 GTCTTAAGATTAAATGAAGTTATTGGTGACAAAATTGGAATGTTCTTCAAGTCAATGGCA 594

Qy 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219

Db 595 ACATTTTTCAGCTGGGTATTAGTAGGATTTACAGCTGGTTGGAAGTAACCCCTTGTGATT 654

Qy 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239

Db 655 TTGCCATCAGTCTCTTCTGGACTGTCAGCTGCTCTGGGCAAGATACTATCTTCA 714

Qy 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259

Db 715 TTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTCTTG 774

Qy 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279

Db 775 GCAGCAATTAGACTGTGATTGCAATTTGGAGGACAAAGAAAGAACTTGAAGGTACAAC 834

Qy 280 LysAsnLeuGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299

Db 835 AAAAATTTAGAAAGAACTAAAAGAAATTTGGGATAAAGAAAGCTATTACAGCAATATTCT 894

Qy 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319

Db 895 ATAGTGTCTGCTTTCTCTGCTGATCTATGATCTTATGCTCTGGGCTTCTGTGATGGGACC 954

Qy 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339

Db 955 ACCTTGGTCTCTCAGGGGAATATTCTATTGGACAAGTACTCACTGTATTCTTCTGTGA 1014

Qy 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359

Db 1015 TTAATTTGGGGCTTTTAGTGTGGACAGGCATCTCCAAAGCATTTGAAGCATTTGCAATGCA 1074

Qy 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379

Db 1075 AGAGGAGCAGCTTATGAAATCTTCAAGATAATTGATAATAAGCCAAAGTATTGACAGCTAT 1134

Qy 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399

Db 1135 TCGAAGAGTGGGCACAAACCCAGATAATATTAAAGGAAATTTGGAATTCAGAAATGTTCC 1194

Qy 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419

Db 1195 TTCAGTTACCCATCTCGAAAAGAAAGTTAGATCTTGAAGGGCTGAACCTGAAGGTGCAG 1254

Qy 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439

Db 1255 ACTGGGCACAGCTGGCCCTGCTTGGAAACAGTGGCTGTGGGAAGAGCACACACAGTCCAG 1314

Qy 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459

Db 1315 CTGATCGCAGAGGCTCTATGACCCACAGAGGGATGGTCAGTGTGATGGCAGAGGATAT 1374

Qy 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479

Db 1375 AGGACCATTAATGTAAGGTTTCTACGGGAAATCATTTGGTGTGTTGAGTCAAGAACCTGTA 1434

Qy 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499

Db 1435 TTCTTTGCCACACGATAGCTGAAAACATTCGCTATGGCCGTGAAATGTCACCATGAT 1494

Qy 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519

Db 1495 GAGATTGAGAAAGCTGTCAAGGAGGCCAATGCCTATGACTTTTATCATGAAATGCCTCAT 1554

QY	520	LysPheAspThrLeuValGlycylArgGlyAlaArgLeuSerGlyGlnLysGlnArg	539
Db	1555	AAATTTGACACCCTGGTTGGAGAGAGGGGCCAGTTGAGTGGCGAGAACGACAGG	1614
QY	540	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	559
Db	1615	ATCGCCATTGACAGTGGCCCTGGTTGCGAACCCCAAGATCCTCTGCTGGATGAGGCCACG	1674
QY	560	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	579
Db	1675	TCAGCCCTTGGACACAGAAAGCGAGCAGTGTTCAGTGGCTCGATGAAGCCAGAAA	1734
QY	580	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	599
Db	1735	GGTCGACCAACCATTTGTGATAGCTCATCTGTTCTACAGTTCTGTAATCTGAGTCA	1794
QY	600	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	619
Db	1795	GCTGGTTTCGATGATGAGTCAATTGTGGAGAAAGGAATCATGATGAATCATGAAGAG	1854
QY	620	LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	639
Db	1855	AAAGGCATTTACTTCAAACTGTGCAAATGCAGACAGCAGGAATGAAGTTGAATTGAA	1914
QY	640	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	659
Db	1915	AATGCAGCTGATGAATCCAAAAGTGAANYTGATGGCTCTGGAATGCTPTCAAATTC	1974
QY	660	GlySerSerLeuLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln	679
Db	1975	AGATCCAGCTCTAATAAGAAAAGATCAACTCGTAGAGTGTCCGTGGATCACAGCCAA	2034
QY	680	AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrp	699
Db	2035	GACAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGCTTCTCTTTGG	2094
QY	700	ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla	719
Db	2095	AGGATTATGAGCTAAATTTAACTGAATGGCCTTATTTGTTGGTGATTTTGTGCC	2154
QY	720	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	739
Db	2155	ATTATAATGGAGCCCTGCACCGCATTTGCAATAATATTTTCAAAAGATTATAGGGTT	2214
QY	740	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	759
Db	2215	TTTACAAGAATTGATGATCCTGAAACAAACACAGAGATAGTAACTGTTTTCACATTTG	2274
QY	760	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	779
Db	2275	TTTCTACCCCTTGGAAATATTCTTTTATACATTTTCTCTCAAGGTTTTCACATTTGCC	2334
QY	780	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	799
Db	2335	AAAGCTGGAGAGATCCTCACCAAGCGCTCCGATACATGTTTTCGATCCATGCTCAGA	2394
QY	800	GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	819
Db	2395	CAGGATGTGATTTGGTTTGTATGACCTAAAAACACCACTGGAGCATTTGATCACCAGCTC	2454
QY	820	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	839
Db	2455	GCCAAATGATGCTGCTCAAGTTAAAGGGCTATAGTTTCCAGGCTTGTCTGAATTA	2514
QY	840	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr	859
Db	2515	AATATAGCAAACTCTGGGACAGGAATAATATATCCCTTCATCTATGTTGGCAACTACA	2574
QY	860	LeuLeuLeuLeuAlaIleValProIleIleAlaIleIleValGluMetLysMet	879
Db	2575	CTGTTACTCTTAGCAATGTACCCATCATGTCAATAGCAGAGTGTCTGAAATGAATG	2634

QY	880	LeuSerGlyGlnAlaLeuLysAspLysLysGlnLeuGluGlyAlaGlyLysIleAlaThr	899
Db	2635	TTGCTGGACAAGCACTGAAAGATAAGAAAGAACTGAAGAGTCTCGGGAAGATCGCTACT	2694
QY	900	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluThr	919
Db	2695	GAAGCAATGAAACACTCCGAACCGTGTTCCTTGNCTCAGAGCAGAAGTTTGAACAT	2754
QY	920	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	939
Db	2755	ATGATGCTCAGAGTTTCAGGTACCATACAGAAACTCTTTGAGGAAGACACACATCTTT	2814
QY	940	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	959
Db	2815	GGAAATTACATTTCTTCACCCAGCGCAATGATGTATTTTCTCATGCTGGATGTTCCGG	2874
QY	960	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe	979
Db	2875	TTTGAGGCTACTTGGTGGCACAATAAATCATGAGCTTTGAGGATGTTCTGTAGTATTT	2934
QY	980	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	999
Db	2935	TCAGCTGTGTCTTTGGTGCATGGCGGTGGGCAAGTCAGTTCATTTGTCCTGACTAT	2994
QY	1000	AlaLysAlaLysValSerAlaAlaHisValIleMetIleGluLysSerProLeuIle	1019
Db	2995	GCCAAAGCCAAATATCAGCAGCCACATCATCATCATGATCATTTGAANAACCCCTTGATT	3054
QY	1020	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1039
Db	3055	GACAGCTACAGCAGCGAAGGCCCTAAATGCCGACACATTTGGAGGAATATGCACATTTGGT	3114
QY	1040	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1059
Db	3115	GAAGTTGTATTCAACTATCCACCCAGCCGACATCCAGTGCCTTCAGGACTGAGGCTG	3174
QY	1060	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1079
Db	3175	GAGGTGAAGAAGGCCAGACGCTGGCTCTGCTGGGCAGCAGTGGCTGTGGAGAGGACACA	3234
QY	1080	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1099
Db	3235	GTGTTCCAGCTCTGGAGCGGTCTACGACCCCTTGGCAGGGAAGTGTCTGTATGGGC	3294
QY	1100	LysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSerGln	1119
Db	3295	AAAGAAATAAAGGCACTGAATGTTCACTGGCTCCGAGCACACCTGGCAGTGTGCCAG	3354
QY	1120	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1139
Db	3355	GAGCCATCTGTTCAGCTGCAGCATGTCTGAGAACAATGGCTATGAGACAACAGCCCG	3414
QY	1140	ValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPheIle	1159
Db	3415	GTGTGTGCACAGGAAGAGATCTGTGAGGCAGCAAAAGGAGGCCACATACATGCTTCATC	3474
QY	1160	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1179
Db	3475	GAGTCACTGCCTTAATAAATATAGCACTAAAGTAGGACACAAGAACTCAGCTCTGTGT	3534
QY	1180	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1199
Db	3535	GGCCAGAAACAACGCATTCCCATACCTCGTCCCTTGTWTAGACAGCCTCATATTTGGCTT	3594
QY	1200	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1219
Db	3595	TTGGATGAAGCCACGTCAGCTCTGTGATACAGAAAGTGAAGAGTTGTCCAAGAACCCCTG	3654
QY	1220	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1239
Db	3655	GACAAAGCCAGAGAGGCCGACCTGCATTTGATTCCTCACCCTGTCTCCACCATCCAG	3714
QY	1240	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1259

QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrRspPheIleMetLysLeuProAsn 519
|||||
Db 1919 GAGATTGAAAGCTGTCAGCAAGCAATGCCTTATGACTTTATCATGAAACATGCCTCAT 1978
QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
|||||
Db 1979 AAATTTGACACCTGGTTGGAGAGAGAGGGCCAGTTGAGTGGTGGCAGAGCAGAGG 2038
QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAlaThr 559
|||||
Db 2039 ATCCCATGTCAGCTGCCTGGTTGCGCAACCCCAAGATCCTCCTGCTGGATGAGGCCAGG 2098
QY 560 SerAlaLeuAspThrGluSerGluAlaValAlaValGlnValAlaLeuAspLysAlaArgLys 579
|||||
Db 2099 TCAGCCTTGGACACAGAAAGCAAGCAGTGGTTCAGGTGGCTCGATAGGCCAGAAA 2158
QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
|||||
Db 2159 GTCGGACCAACCATTTGTATGATGCTATCGTTTCTACAGTTCTGTAATGCTGACGTCATC 2218
QY 600 AlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
|||||
Db 2219 GCTGTTTTCGATGATGGAGTCATTTGGAGAAAGGAATCATGATGAATCATGAAGAG 2278
QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
|||||
Db 2279 AAAGGCATTTACTTCAAACTTGTCAATGTCACAGCAGCAAGAAATGAAGTTGAATAGAA 2338
QY 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
|||||
Db 2339 AATGCAGCTGATGAATCCAAAGTGAATTTGATGCTTGGAAATGCTTCAAAATGATTTCA 2398
QY 660 GlySerSerLeuLysArgSerThrArgSerIleHisAlaProGlnGlyGln 679
|||||
Db 2399 AGATCCAGCTTAATAAGAAAGATCAACTCGCTAGGAGTGTCCGTCGATCACAGCCCA 2458
QY 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrp 699
|||||
Db 2459 GACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTTGG 2518
QY 700 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719
|||||
Db 2519 AGGATTATGAAGCTAAATTTAACTGAATGGCTTATTTTGTGTGGTGTATTTTGTGCC 2578
QY 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
|||||
Db 2579 ATTATAAATGGAGGCTGCAACAGCATTGCAATAATATTTTCAAGATATATAGGGGTT 2638
QY 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
|||||
Db 2639 TTTACAAGATTGATGATCCTGGAACAAACACAGACAGATAGTAATGTTTTCATCTATTG 2698
QY 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
|||||
Db 2699 TTTTCAGCCTTGGAAATTTATTTTATTTATTTTTCCTTTCAGGCTTCACATTGGC 2758
QY 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
|||||
Db 2759 AAAGCTGGAGAGATCCTCACCAGCGGCTCGATACATGTTTTCGATCCATGCTCAGA 2818
QY 800 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
|||||
Db 2819 CAGGATGTGAGTTGGTTGATGACCTTAAACACACCACTGGAGCATTTGACTACAGGCTC 2878
QY 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
|||||
Db 2879 GCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTTCCAGGCTTGCCTGTAATTTCCAG 2938
QY 840 AsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThr 859
|||||
Db 2939 AATATAGCAATCTTGGACAGGAATAATATATCTCTCATCTATGTTGGCAACTAACA 2998
QY 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
|||||

Db 2999 CTGTGTACTCTTAGCAATTTGTACCCATCATTTGCAATACAGGAGTGTGTGAATGAAATG 3058
QY 880 LeuSerGlyGlnAlaLeuLysAspLysGlnLeuGluGlyAlaGlyLysIleAlaThr 899
|||||
Db 3059 TTGCTGTGACAAAGCAGCTGAAAGATAAGAAAGAACTAGAGGTGCTGGAGATCGCTACT 3118
QY 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
|||||
Db 3119 GAAGCAATAGAAAACCTCCGACCGTGTCTTTGACTCAGGAGCAGAGTTTGAACAT 3178
QY 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
|||||
Db 3179 ATGTATGCTCAGAGTTTTCAGGTACCATACAGAAACTCTTTGAGGAAACACACATCTTT 3238
QY 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
|||||
Db 3239 GGAATTACATTTTCCATCCAGGCAATGATGATATTTTCTATGCTGGATGTTTCCGG 3298
QY 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 979
|||||
Db 3299 TTTGGAGCCTACTTGGTGGCACATAAACTCATGAGCTTTGAGGATGCTTCTGCTGACTAT 3358
QY 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
|||||
Db 3359 TCAGCTGTTGCTTGTGTCATGGCCGTGGGGCAAGTCAGTTTCATTTGCTCCTGACTAT 3418
QY 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
|||||
Db 3419 GCCAAAGCCAAATATACAGAGCCCATCATCATGATCATTTGAAAAAACCCCTTTGAT 3478
QY 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
|||||
Db 3479 GACAGCTACAGCACCGAAGGCGCTTAATGCCGAACACATTTGAAGGAATGTCACATTTGGT 3538
QY 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
|||||
Db 3539 GAAGTTGTTATTCACATATCCACCCGACCGGACATCCAGTGTTCAGGGACTGAGCGTG 3598
QY 1060 GluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
|||||
Db 3599 GAGTGAAGAAAGGGCCAGACGCTGCTCTGTGGGCGACAGTGGCTGTGGGAGAGCACA 3658
QY 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
|||||
Db 3659 GTGGTCCAGCTCCTGGAGCGGTCTACAGCCCTTGGCAGGAAAGTGTCTGTTGATGGC 3718
QY 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
|||||
Db 3719 AAAGAAATAAAGCGACTGAATGTTTCAGTGGCTCCGAGCACACCTGGGCATCGTGTCCAG 3778
QY 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
|||||
Db 3779 GAGCCATCCTGTTTGTGCTGAGCATTTGCTGAGAACATTTGCTATGAGACAAACAGCCGG 3838
QY 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1159
|||||
Db 3839 GTGGTGTACAGGAAGAGATCGTGGGGCAGCAAGAGGCGCAACATACATGCTTCATC 3898
QY 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
|||||
Db 3899 GAGTCACCTGCCTTAATAATATAGCATAAGTAGGAGACAAAGAACTCAGCTCTCTGGT 3958
QY 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
|||||
Db 3959 GCCAGAAACCAACCGCATTTGCCATAGCTCGTCCCTTTTATAGACAGGCTCATATTTGCTT 4018
QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
|||||
Db 4019 TTGGATGAAGCCAGCTCAGCTGGATACAGAAAGTGAAGAGTTGTCCAAAGGCCCTG 4078
QY 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
|||||

Db 1859 TTGTTGGCCACCATAGCTGAAACATTCGCTATGCGCGTGAATGTCACCATGGAT 1918
 QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaThrPheMetLysLeuProAsn 519
 Db 1919 GAGATTGAGAAGCTGTCAAGGAAGCCCAATGCCCTATGACTTTATCATGAAACTGCCCTCAT 1978
 QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
 Db 1979 AAATTTGACACCCCTGGTTGGAGAGAGAGGGGCCAGTTGAGTGGTGGCAGAGACAGAGG 2038
 QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
 Db 2039 ATCGCCCATTTGACCTGGCCCTGGTTGCGAACCCCAAGATCCTCCTGCTGGATGAGGCCACG 2098
 QY 560 SerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArgLys 579
 Db 2099 TCAGCCTTGGACACAGAAAGCAAGCAGTGGTTTCAGGTGGCTCTGGATAAGGCCAGAAA 2158
 QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
 Db 2159 GGTGGACCATTTGATAGCTATCGTTTGTCTACAGTTCTGTAATGCTGACGTCATC 2218
 QY 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
 Db 2219 GCTGGTTTCGATGATGAGTCATTTGTGGAGAAAGAAATCATGATCAACTCATGAAGAG 2278
 QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
 Db 2279 AAAGGCATTTACTTCAAACTTGTCCACNATGCACAGCAGCAGGAAATGAAGTTGAATAGAA 2338
 QY 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
 Db 2339 AATCAGCTGATGAATCAAAAGTGAATTTGATGCTTGGAAATGCTTCAATGATCA 2398
 QY 660 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
 Db 2399 AGATCCAGTCTAATAGAAAGATCACTCGTAGAGTGTCCGTGGATCAACAGCCCAA 2458
 QY 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheThrP 699
 Db 2459 GACAGAAAGCTTAGTACCAAGAGGCTCGGATGAAGATATACCTCCAGTTTCCCTTTGG 2518
 QY 700 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719
 Db 2519 AGGATTATGAAGCTAAATTAAGTAAATGGCCTTATTTGTGTTGGTATTTTGTGCC 2578
 QY 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
 Db 2579 ATTTAAATGGAGGCTGCAACAGCATTTGCAATAATATTTTCAAGATATATAGGGGTT 2638
 QY 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
 Db 2639 TTTTACAGAATGATGATCCTCAACAAACACACAGATAGTAATCTGTTTTCCTACTATGT 2698
 QY 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
 Db 2699 TTTTCTAGCCCTTGGAAATATTTCTTTTATTTATATTTTCTTCAGGTTTTCACATTTGGC 2758
 QY 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
 Db 2759 AAAGCTGGAGAGATCCTCACCAGGGCTCCGATACATGAGTGGTTTCCGATCATGCTCAGA 2818
 QY 800 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuLeuThrThrArgLeu 819
 Db 2819 CAGGATGTGAGTTGGTTTGTATGACCTTAAACACACCTGGAGCATTGACTACAGGCTC 2878
 QY 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
 Db 2879 GCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGTTTCCAGGCTTGTGTAATTTACCCAG 2938
 QY 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 859
 Db 8409 TTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAAGAGCCCTG 4078

Db 2939 AATATAGCAAAATCTTTGGGACAGGAATAATATATCTTCATCTATGCTTGGCAACTAACA 2998
 QY 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
 Db 2999 CTGTTACTCTTAGCAATTTGTACCCATCATTCATACAGAGAGTTGTTGAATGAAGATG 3058
 QY 880 LeuSerGlyGlnAlaLeuLysAspLysGluLeuGluAlaGlyLysIleAlaThr 899
 Db 3059 TTGCTGACAAAGCACTGAAAGATAAGAAAGACTAGAAAGTGTGGAAAGATCGCTACT 3118
 QY 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
 Db 3119 GAAGCAATAGAAACTTCCGAAACGTTGTTTCTTTGACTCAGGAGCAGAGTTTGAACAT 3178
 QY 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
 Db 3179 ATGTATGCTCAGATTTGCAGGTACCATACAGAACTCTTTTGAGGAAAGCAGACATCTTT 3238
 QY 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
 Db 3239 GGAATTCATTTTCTTCAACCCAGGCAATGATGATTTTCTTATCTGATGCTGATGTTCCGG 3298
 QY 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 979
 Db 3299 TTTGGAGCCTACTTGGTGGCACAATAAATCATGAGCTTTGAGGATGTTCTGTTAGTATTT 3358
 QY 980 SerAlaIleValPheGlyAlaMetAlaValGlnValSerSerPheAlaProAspTyr 999
 Db 3359 TCAGCTGTGTTGTTGTCATGGCGTGGGCAAGTCAAGTTCATTTCTCTGCTGACATAT 3418
 QY 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
 Db 3419 GCCAAAGCCAAATATCAGAGCCCATCATCATCATCATCATCAATGATCAATGAAAGAA 3478
 QY 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
 Db 3479 GACAGCTACAGCAGCGGAAGGCCTAATGCCGACACACATTCGAAAGAAATGTCACATTTGT 3538
 QY 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
 Db 3539 GAAGTTGTATTTCAACTATCCACCCAGCGCATCCCAAGTCTTCAGGAGCTGAGCCTG 3598
 QY 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
 Db 3599 GAGGTGAAGAAAGGCCAGAGCTGGCTGTGGGCGCAGCAGTGTGTGGGAAAGAGCACA 3658
 QY 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
 Db 3659 GTGGTCCAGCTCTCGAGCGGTTCTACGACCCCTTGGCAGGAAAGTCTGCTGTGATGGC 3718
 QY 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
 Db 3719 AAAGAAATAAAGCGACTGAATTTCAAGTGGCTCCGAGCAGCACCATATACATGCTTCATC 3778
 QY 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
 Db 3779 GAGCCCATCTGTTTGTACTGCAGCATTTCTGAGAATCTGCTATGGAGAACAGACGCCGG 3838
 QY 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1159
 Db 3839 GTGGTGTACAGAAAGAGATCGTAGGGCAGCAAGAGGAGGCCAATACATGCTTCATC 3898
 QY 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
 Db 3899 GAGTCACCTGCTTAATAATATAGCACTAAAGTAGGAGACAAAGAACTAGCTCTCTGTGT 3958
 QY 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
 Db 3959 GSCCAAGAAACACGATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4018
 QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
 Db 4019 TTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAAGAGCCCTG 4078

Thu Nov 7 09:13:29 2002

Thu Nov 7 09:13:29 2002

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:36:49 ; Search time 301.41 Seconds
(without alignments)
7291.243 Million cell updates/sec

Title: US-09-672-725C-4
Perfect score: 6467
Sequence: 1 MDPEGRKGSAEKNFKMGK.....LLAOKGIYFMSISVQAGAKR 1280

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlh
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725.ecgn_1.1.0 -runat_04112002.124338.17877 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -YGAPOP=6 -YGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_032802:*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	6467	100.0	4317	22	AAD03489 Dog P-glycoprotein
2	6425.5	99.4	4279	22	AAD03504 Dog P-glycoprotein
3	6425.5	99.4	4279	22	AAD03505 Dog P-glycoprotein
4	6425.5	99.4	4279	22	AAD03506 Dog P-glycoprotein
5	6420.5	99.3	4279	22	AAD03488 Dog P-glycoprotein
6	5859.5	90.6	3860	21	AAD49332 Human wild-type mu
7	5859.5	90.6	3860	21	AAD49332 Human BCRP DNA rel
8	5855.5	90.5	4186	22	AAF86127 Cynomologous monke
9	5853	90.5	4195	22	AAF86128 Human G185V mutant
10	5850.5	90.5	3860	21	AAD49333 Human BCRP DNA rel
11	5850.5	90.5	3860	24	ABA94366 Human intestine ce
12	5850.5	90.5	4349	22	AHH57442 Human ATP binding
13	5850.5	90.5	4646	21	AAZ94738 Hybrid vector pSF-
14	5850.5	90.5	6505	17	AAZ11394 Retroviral M4 mdr-
15	5850.5	90.5	8630	21	AAZ24041 Retroviral vector
16	5850.5	90.5	8630	21	AAZ24041 Sequence of human
17	5847.5	90.4	4669	8	AAZ24042 Multidrug Resistan
18	5844.5	90.4	4378	11	AAQ04522 Human MDR-1 DNA.
19	5843	90.4	3988	21	AAZ88973 Human multidrug re
20	5837.5	90.3	4646	15	AAQ72872 Sequence of human
21	5833.5	90.2	4669	14	AAV32645 Human P-glycoprote
22	5823.5	90.0	4659	19	AAV32645 Mutated human P-gl
23	5809	89.8	4264	19	AAV66533 Mutated human P-gl
24	5809	89.8	4264	19	AAV66533 Murine multidrug r
25	5638.5	87.2	4788	21	AAZ49335 Rat multidrug resi
26	5638.5	87.2	4788	24	ABA94368 Rat multidrug resi
27	5621.5	86.9	4425	21	AAZ52047 Rat multidrug resi
28	5611.5	86.8	4369	21	AAZ52047 Rat multidrug resi
29	5304.5	82.0	4189	24	ABA94367 Rat mdr1b2 (multis
30	5304.5	82.0	4313	14	AAQ38950 Rat mdr1b2 (multis
31	5304.5	82.0	4313	14	AAQ38950 Human ATP binding
32	5296	81.9	4233	21	AAZ88974 Human MDR-3 DNA.
33	5296	81.9	4233	22	AAZ88974 Multidrug-resistan
34	4898	75.7	3924	21	AAZ88974 Multidrug-resistan
35	4898	75.7	3924	21	AAZ88974 Human ATP binding
36	3615	55.9	2726	15	AAQ70907 Multidrug-resistan
37	3615	55.9	2726	15	AAQ70907 Multidrug-resistan
38	3615	55.9	2726	18	AAZ43322 Human ATP binding
39	3392	50.9	4776	21	AAZ94744 H. contortus PGP-A
40	2810	43.5	4175	20	AAV69392 Drosophila melanog
41	2685.5	41.5	3942	23	ABL14175 Drosophila melanog
42	2554	39.5	4157	23	ABL06835 cDNA encoding the
43	2511.5	38.8	4047	20	AAZ60201 Drosophila melanog
44	2491	38.5	4111	23	ABL16205 cDNA encoding mult
45	2482	38.4	4002	20	AAZ85823

ALIGNMENTS

RESULT 1

AAD03489
ID AAD03489 standard; cDNA; 4317 BP.
XX
XX AAD03489;
AC
DT 13-JUN-2001 (first entry)
XX
DE Dog P-glycoprotein (PGP) cDNA #2.
XX
KW Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW drug bioavailability; transgenic animal; genetic model; ss.
XX
XX Canis familiaris.
XX
FH Key Location/Qualifiers
CDS 70..3912
FT /*tag= a

FT /product= "Dog P-glycoprotein (PGP) #2"

PN WO200123540-A2.

XX 05-APR-2001.

PD 28-SEP-2000; 2000WO-US26767.

XX 28-SEP-1999; 99US-0156510.

PR (GENT-) GENTEST CORP.

XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;

PI WPI; 2001-235373/24.

XX P-PSDB; AAE00304.

DR New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful

XX for determining the bioavailability of drugs and for screening for dog

PT PGP inhibitors -

XX Claim 1; Page 66-72; 11pp; English.

XX The invention relates to dog P-glycoprotein (PGP) also referred

CC as multidrug transporter (MDR1) and nucleic acids encoding them.

CC The invention also includes fragments and biologically functional

CC variants of dog P-glycoprotein. PGP and their nucleic acids are

CC useful for determining the bioavailability of drugs and for

CC screening PGP inhibitors. They are useful for the diagnosis and

CC treatment of conditions characterised by PGP activity, by

CC reducing or increasing PGP activity in a cell. PGP nucleic acids

CC are used as oligonucleotide probes. Complements of PGP nucleic

CC acids are useful as antisense oligonucleotides, to induce a PGP

CC 'knockout' phenotype. They are used to prepare a non-human

CC transgenic animal, which are valuable as genetic models for

CC human diseases.

CC The present sequence is dog P-glycoprotein (PGP) cDNA. The

CC PGP enzyme functions as an efflux pump exporting small molecules

CC across the cell membrane. This enzyme is a member of the ABC

CC transporter family.

XX Sequence 4317 BP; 1293 A; 844 C; 1019 G; 1161 T; 0 other;

50

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0	6467.00	4317	1280
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-09-672-725C-4 (1-1280) x AND03489 (1-4317)

QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20

Db 70 ATGGATCTCTGAAGGAGCGCTAAGGGAGTGCAGAGAAGAACTTCTGGAAATGGGCAA 129

QY 21 LysSerLysLysGluLysGluLysLysProThrValSerThrPheAlaMetPheArg 40

Db 130 AAAGTAAAG 189

QY 41 TyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleHis 60

Db 190 TATTCAATTTGGCTTGATAGGTGTATATGTTGGTGGGACAATGGCTGCCATCATCCAT 249

QY 61 GlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAsn 80

Db 250 GGAGCTGCACCTCCCTCATGATGCTGGTTTTTGGAAACATGACAGATAGCTTTGCAAT 309

QY 81 AlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsn 100

Db 310 GCAGGAATTTCAAGNAACAAACTTTTCCAGTTATATAATTAATGAAGTATTACGAACAAT 369

QY 101 ThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrSer 120

Db 370 ACACACATTTTCATCAACCATCTGGAGGAGAAATGACACGATATGCTATTATTACAGT 429

QY 121 GlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeuAla 140

Db 430 GGGATCGTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 489

QY 141 AlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGlu 160

Db 490 GCAGAGACAGATACTCAAAATTAGAAAACAATTTTTCATGCTATCATCGCAGAGAG 549

QY 161 IleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAspVal 180

Db 550 ATTGGCTGGTTGACGTGACGTGACGTGACGTGACGTGACGTGACGTGACGTGACGTG 609

QY 181 SerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAlaThr 200

Db 610 TCCAAAATCAATGAAGGAATTTGGCGACAAAGTTGGAATGTTCTTCAATCAATAGCAAC 669

QY 201 PhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIleLeu 220

Db 670 TTTTTCACCGGTTTATAGTGGGTTTACACCTGGTTGGAAGCTAACCCCTTGTGATTTG 729

QY 221 AlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPhe 240

Db 730 GCCATCAGCCCTGTTCTTGGACTTTTCAGCCGCATCTGGCAAGATACTATCTCATTT 789

QY 241 ThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeuAla 260

Db 790 ACTGATAAGAACTCTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAAGAAGTCTTAG 849

QY 261 AlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyrAsnLys 280

Db 850 GCAATCAGAACTGTGATTGCTTTGGAGACAAAAGAAAGAACTTGAAGAGTGACACAAA 909

QY 301 GlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTrpGlyThrSer 320

Db 970 GGTGGCGCTTCTTATTGATCATGATCATGCTGCTGGCTTCTGGTATGGGACCTCC 1029

QY 321 LeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerValLeu 340

Db 1030 TTGGTCTCTCCAGTGAATATATTTGGACAGGTAATCTCCTCTCTCTCTCTCTCTCTCT 1089

QY 341 IleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAlaArg 360

Db 1090 ATTGGGCTTTTAGTATTGGACAGGCATCCCAAGCATTTGAAGCATTTGCAACGCAAG 1149

QY 361 GlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyrSer 380

Db 1150 GGAGCAGCTTATGAAATCTTCAAGATAATTGACAAATAAACCAAGCATTTGACAGCTATT 1209

QY 381 LysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPhe 400

Db 1210 AAGAGTGGACATTAACAGATATATTAAGGAAATTTGGAATTTCAAAAATGTTCACTTC 1269

QY 401 SerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSer 420

Db 1270 AGTTACCTTCTCGAAAAGAAAGTAAAGATCTTAAAGGGTCTCAACCTGAAGGTTACAG 1329

QY 421 GlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeu 440

Db 1330 GGGCAGACAGTGGCGCTGGTTGGGAACAGTGGCTGGGGAAGACGACGACCGTGACGCTG 1389

QY 441 MetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArg 460

Db 1390 ATGCAGAGGCTCTATGACCCACAGATGGCATGGCTGTGTTATGATGGACAGGACATAG 1449

QY 461 ThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeu 480

|||||
Db 1450 ACCATAAATGTAAGGCATCTTCGGAAATTAATGCTGCTGCTGAGTACGAGGCTGTGTTG 1509
QY PheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAspGlu 500
|||||
Db 1510 TTTCGCCACCGAGTAGCTGAAACATTCGCTATGCGCGGAAATGTCACCATGGATGAG 1569
QY IleGluIysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsnLys 520
Db 1570 ATTGAGAAAGCTGTTAAGGAGCCATGCCATGATTTATCATGAACTACCTAATAAA 1629
QY PheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArgIle 540
|||||
Db 1630 TTTCGACACTCTGTTGGAGAGAGGCGCGCTGAGTGGTGCACAGAAACAGAGAAATC 1689
QY AlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThrSer 560
Db 1690 GCCATTGCTCGGCGCTGGTTCGCAACCCCAAGATTTCTGCTGGATGAGGCAACGTCA 1749
QY AlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLysGly 580
Db 1750 GCTCTGACACTGAAAGTGAAGCAGTGGTTTCAGTGGCCCTGGATAAGGCCAGAAAGCC 1809
QY ArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla 600
Db 1810 CGGACTACCATGTGTAGTACATCGTTGCTCTACAGTTTCGTAATGCCGATGTCATTGCT 1869
QY GlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGluLys 620
Db 1870 GGTTTTGATGATGGAGTCATTGTGGAGAAAGGAATCATGATGAACATCATGAAAGAGAAG 1929
QY GlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGluAsn 640
Db 1930 GGCATTACTTCAAACTTGTCAATGCAGACAGAGGAATGAAATGAGTTAGTAATAAT 1989
QY AlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSerGly 660
Db 1990 GCCACTGGTGAATCCAAAAGTGAAGTGATGCTTGGAAATGTCTCCAAAAGATTCAAGG 2049
QY SerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAsp 680
Db 2050 TCCAGTTTAATAAAGAAGATCAACTCGCAGGAGTATACATGCACACAGGCCAAGAC 2109
QY ArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTyrArg 700
Db 2110 AGAAACCTGGTACAAAAGAGACTGAATGAGAAATGTCTCTCAGTTTCCTCTGGAGG 2169
QY IleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAlaIle 720
Db 2170 ATTCTGAAGCTGAACCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 2229
QY IleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePhe 740
Db 2230 ATAAACGGAGGCTGCAACACAGCATTTTCAATAATATTTTCAAGGATATAGGATCTTT 2289
QY ThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe 760
Db 2290 ACCCGAGATGAGATCCTGAACAAACAGCAGAAATAGTAACATGTTTCTGTATTGTTT 2349
QY LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys 780
Db 2350 CTAGTCTTGGAAATTAATTTCTTTTATTACATTTTCTCCAGGCTTCACATTGGCNA 2409
QY AlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGln 800
Db 2410 GCTGGGAGATCCTCACTAAGCGGCTTGCATACATGTTTTCAGATCCATGCTGAGACAG 2469
QY AspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla 820
Db 2470 GATGTCAGCTGGTTGATGACCTTAACACACCACCTGGAGCATTGAAACAGGCTGGC 2529
QY AsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn 840
|||||

Db 2530 AATGATCGGCTCAAGTATAAGGGGCTATAGTTTCCAGGCTTGTGTCATTACCCAGAA 2589
QY IleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTyrGlnLeuThrLeu 860
|||||
Db 2590 ATAGCAATCTTGGACAGCATTAATTAATATCTTAATCTATGTTGGCAATTAACACATT 2649
QY LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMetLeu 880
Db 2650 TTACTCTTAGCAATGTACCCATCATTTGCAATAGCAGGAGTTCTTGAATGAAATGTTG 2709
QY SerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAlaThrGlu 900
Db 2710 TCTGCACAGCCTGAAGATAGAAGAGCTAGAAGAGCTGGGAAGATTGCTACAGAA 2769
QY AlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyrMet 920
Db 2770 GCCATCGAAAACCTCCGAACGTGTTGTTCTTCTGACTCGGAGCAGAAAGTTTGAATACATG 2829
QY TyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGly 940
Db 2830 TATGCACAGAGTTTGCAGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTTCGGG 2889
QY ValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPhe 960
Db 2890 GTCTCATTTCTATCACCCAGGCAATGATGATTTTCTATGCTGGCTGTTCGGGTTT 2949
QY GlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSer 980
Db 2950 GGTGCTACTTGTGGCAATGAGTTCATCACTTTCAGGATGTTCTTTGGTATTCATCA 3009
QY AlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyrAla 1000
Db 3010 GCTATTGCTTTGGTGCCATGGCAGTGGGGAGGTCAGTTCAATTTGCTCCTGACTATGCC 3069
QY LysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIleAsp 1020
Db 3070 AAAAGCAAGTATACAGACGCCAGCTCATCATGATCATTTGAAAAAAGCCCTCTGATTGAC 3129
QY SerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGlu 1040
Db 3130 AGCTACAGCCCTCACGGCTCAAGCCAAATACGTTTGAAGGAAATGTGACATTTAATGAG 3189
QY ValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeuGlu 1060
Db 3190 GTGCTGTTCAACTATCCCACTCGACACACATCCCGCTGCTCCAGGGCTCAGGCTCGAG 3249
QY ValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrVal 1080
Db 3250 GTGAAGAGGGCCAGACGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGACACAGT 3309
QY ValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGlyLys 1100
Db 3310 GTTCAGTCTCTAGAGCGCTTCTATACCCCTTGGCTGCTCAGTGTCAATTTGATGGCAA 3369
QY GluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSerGlnGlu 1120
Db 3370 GAGATAAAGCACTCAATGTCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAGAG 3429
QY ProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgVal 1140
Db 3430 CCCATCTCTTTGAGCTGAGCATTTGCCGAGAACATTCCTATGAGACACACAGCCGGTCT 3489
QY ValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIleGlu 1160
Db 3490 GTATCACATGAAGAGATTATCGAGGACCGAAGGAGGCGCAACATACACCACTTCATCGAG 3549
QY ThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGly 1180
Db 3550 ACACCTCCCTGAGAAATACACACAGAGTAGGAGACAAAGAAACCCAGCTCTCTGTGGGC 3609
QY GlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeu 1200
Db 3610 CAGAAACAGCCGATGGCATAGCTCGCGCTCTTGTGTAGACAGGCTCATATTTTGTCTTTG 3669

QY 1201 AspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGlnAlaLeuAsp 1220
|||||
Db 3670 GATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAAGAGCCCTGGAC 3729
QY 1221 LysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsn 1240
|||||
Db 3730 AAAGCCAGAGAGCGCCACCTGCAFTGTGATCGCCACCGCTTGTCCACCATCCAGAT 3789
QY 1241 AlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGlnGln 1260
Db 3790 GCAGATTATAGTGTGTTTCAGATGCGAAAGTCAAGGACATGGCCACACATCAACAG 3849
QY 1261 LeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLysArg 1280
|||||
Db 3850 CTGCTGCTCAGAAAGGCATCTATTTTTCATGATCAGTGTCCAGGCTGGAGCAAGCCG 3909

RESULT 2
AAD03504
ID AAD03504 standard; cDNA; 4279 BP.
AC AAD03504;
XX 13-JUN-2001 (first entry)
DT
XX
DE Dog P-glycoprotein (PGP) allelic variant (Genotype A) cDNA.
XX
KW Dog; P-glycoprotein allelic variant; multidrug transporter; MDR1;
KW drug bioavailability; transgenic animal; genetic model; ss.
XX
OS Canis familiaris.

FH Key Location/Qualifiers
FT CDS 17..3862
FT /*tag= a
FT /product= "Dog P-glycoprotein (PGP) allelic variant
FT (Genotype A) protein"
FT allele replace (607, C)
FT /*tag= b

XX WO200123540-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26767.
XX
XX 28-SEP-1999; 99US-0156510.
XX
XX (GENT-) GENTEST CORP.
XX
XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX
XX WPI; 2001-235373/24.
XX
XX P-PSDB; AAE00308.
XX
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX for determining the bioavailability of drugs and for screening for dog
XX PGP inhibitors -
XX
XX Claim 9; Page 85-90; 111pp; English.

XX The invention relates to dog P-glycoprotein (PGP) also referred
XX as multidrug transporter (MDR1) and nucleic acids encoding them.
XX The invention also includes fragments and biologically functional
XX variants of dog P-glycoprotein. PGP and their nucleic acids are
XX useful for determining the bioavailability of drugs and for
XX screening PGP inhibitors. They are useful for the diagnosis and
XX treatment of conditions characterised by PGP activity, by
XX reducing or increasing PGP activity in a cell. PGP nucleic acids
XX are used as oligonucleotide probes. Complements of PGP nucleic
XX acids are useful as antisense oligonucleotides, to induce a PGP
XX 'knockout' phenotype. They are used to prepare a non-human
XX transgenic animal, which are valuable as genetic models for

CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) allelic variant
CC (Genotype A) cDNA. The PGP enzyme functions as an efflux pump
CC exporting small molecules across the cell membrane. This enzyme
CC is a member of the ABC transporter family.
XX
SQ Sequence 4279 BP; 1295 A; 833 C; 1008 G; 1143 T; 0 other;
Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6425.50 Matches: 1273
Percent Similarity: 99.69% Conservative: 4
Best Local Similarity: 99.38% Mismatches: 3
Query Match: 99.36% Indels: 1
DB: 22 Gaps: 1
US-09-672-725C-4 (1-1280) x AAD03504 (1-4279)

QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTyrLysMetGlyLys 20
|||||
Db 17 ATGGATCCTGAAGGAGCGCGTAAGGGAGTGCAGAGAAAGAACTTCTGGAAATGGGCAAA 76
QY 21 LysSerLysLys---GluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
|||||
Db 77 AAAAGTAAAAAATGACAGAGAAAGAAACCACTGTCCAGCACGTTTGCATGTGT 136
QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 59
|||||
Db 137 CGCTATTCAAATTTGGCTTGATAGTTGTATATTTGGTGGGACAAATGGCTGCCATCATC 196
QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
|||||
Db 197 CATGGAGCTGCACCTCCCTCTCATGATGCTGGTTTTTGGAAACATGACAGATAGCTTTCGA 256
QY 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
|||||
Db 257 ATGCGAGGAATTTCAAGAAACAAACTTTTCCAGTTATATAATTAATGAAGTATTACGAAC 316
QY 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrTyrAlaTyrTyrTyr 119
|||||
Db 317 AATACACAAACATTTTCATCAACCATCTGGAGGAGAAATGACCCATATGCTATATTATAC 376
QY 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
|||||
Db 377 AGTGGGATCGGTGCTGGCGTGTGCTGTACATCCAGGTTTCATTCCTGTCCTG 436
QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
|||||
Db 437 GCAGCAGGAAGACAGATACATACTCAAAATTTGAAAACAATTTTTTCATGCTATCATGCGACAG 496
QY 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
|||||
Db 497 GAGATTGGCTGGTTTGCATGACGTTGGGAGCTTAACACCGGCTCACAGACGAT 556
QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
|||||
Db 557 GTCTCCAAATCAATGAAGGAATGGCGACAAATTTGSAATTTCTTTCATCAATCAGATAGCA 616
QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
|||||
Db 617 ACATTTTTCACCGGTTTTATAGTGGGGTTTACACGTGGTGGAGCTAACCCCTGTGTGATT 676
QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
Db 677 TTGGCCATCAGCCCTGTCTTGGACTTTCAGCCGCTTTCAGCCGCTTTCAGCCGCTTTCAG 736
QY 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
Db 737 TTTTACTGATAAAGAACTCTTTGGGCTATGCAAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 796
QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
Db 797 GCAGCAATCAGAACTGTGATGTGCTTTGGAGGACAAAAAGAAAGAACTTTGAAAGGTACAAC 856

QY	280	LysAsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIlePheAlaAsnIleSer	299
DB	857	AAAAATTAGAAAGAGCTAAAGGAATTGGGATAAAGAAAGCTATCACGCCAACATTCT	916
QY	300	IleGlyAlaAlaPheLeuLeuIleIleTyAlaSerTyAlaLeuAlaPheTrpTyGlyThr	319
DB	917	ATTGGTGCCGCTCTCTTAATGATCATGTGCATCATATGCTCTGGCTTCTTGTTATGGGACC	976
QY	320	SerLeuValLeuSerSerGluTyThrIleGlyGlnValLeuThrValPhePheSerVal	339
DB	977	TCCTTGGTCTCTCCAGTGAATATCTATTGGACAGTAGTACTCAGTGTCTCTCTTCTGTGA	1036
QY	340	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	359
DB	1037	TTAATTTGGGGCTTTTACTATTGGACAGCATCCCAAGCATTTGAAGCATTTGCAACGCA	1096
QY	360	ArgGlyAlaAlaTyTGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	379
DB	1097	AGAGGAGCAGCTTATGAATCTTCAGATAATTGACATAAACCAGCATTTGACAGCAT	1156
QY	380	SerLysSerGlyHisLysLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	399
DB	1157	TCGAAGAGTGCACATAAACACAGATAATATAAGGGAATAATTGGAATTCAAAAATGTTCCAC	1216
QY	400	PheSerTyProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	419
DB	1217	TTCAGTTACCTCTCTGAAAGAAGATTAGATCTTTAAAGGGTCTCACTGAAGGTTTCAG	1276
QY	420	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln	439
DB	1277	AGTGGGCAGACAGTGGCGCTGGTGGGAACAGTGTCTCGGGAAGACGACGCGTCGAC	1336
QY	440	LeuMetGlnArgLeuTyArgProThrAspGlyMetValCysIleAspGlyGlnAspIle	459
DB	1337	CTGATGCAGAGGCTCTATGACCCACAGATGGCATGGTCTGTATGTGTGACAGGACATT	1396
QY	460	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	479
DB	1397	AGGACCATAATTAAGGCATCTTCGGGAAATTACTGCTGTGGTGACTCAGGAGCCTGTG	1456
QY	480	LeuPheAlaThrThrIleAlaGluAsnIleArgTyArgGlyArgGluAsnValThrMetAsp	499
DB	1457	TTGTTTGCACACAGTAGCTGAACAATTCGCTATGGCCGCGAAAATGTCAACATGGAT	1516
QY	500	GluIleGluLysAlaValLysGluAlaAsnAlaTyArgPheIleMetLysLeuProAsn	519
DB	1517	GAGATTGAGAAAGCTGTTAAGGAAGCAATGCCTATGATTTATCATGAACATACCTAAT	1576
QY	520	LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg	539
DB	1577	AAATTGCACACTCTGTTTGAGAGAGGGCCACGCTGATGTGTGGACAGAAACAGAGA	1636
QY	540	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	559
DB	1637	ATCGCCATTCTCGGSCCCTGGTTGCAACCCCAAGATTCTTCTGTGTGATGAGGCAAGC	1696
QY	560	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	579
DB	1697	TCAGCTCTGGACACTGAAGTAGTGAAGCAGTGGTTTCAGCTGGCCCTGGATAAGGCCAGAAA	1756
QY	580	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	599
DB	1757	GGCCGGACTACCATTTGTGATAGCTCAUCGTTTGTCTACAGTTCTGTAATCCGATGTCAATT	1816
QY	600	AlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	619
DB	1817	GCTGGTTTGTGATCGGAGTCAATTCTGGAGAAAGGAAATCATCATCACTCATGAAGAG	1876
QY	620	LysGlyIleTyPheLysLeuValThrMetGlnThrArgGlyAsnGluIleLeuGlu	639
DB	1877	AAGGGCATTTACTTCAAACTTGTCCAAATGCACAGAGGAATGAATTCAGTTAGAA	1936

QY	640	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	659
Db	1937	AATGCCACTGGTGAATCCAAAGGTGAAGTGCCTTGGAAATGTCCTCAAAGATTCA	1996
QY	660	GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln	679
Db	1997	GGGTCCAGTTTAATAAAAGAGATCAACTCGCAGAGATATACATGCACCAAGGCCAA	2056
QY	680	AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrp	699
Db	2057	GACAGAAAGCTTGGTACAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCCCTTCGG	2116
QY	700	ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla	719
Db	2117	AGGATTCGAGCTGAAGTCACTCACTGAATGGCCTATTATTTGTGGTGGTATATTTGTGCT	2176
QY	720	IleIleAsnGlyLysLeuGlnProAlaPheSerIleIlePheSerArgIleIleLeuIle	739
Db	2177	ATTATAAACGGAGGCCCTGCAACAGGACATTTTCAATAATATTTCAAGGATTATAGGATC	2236
QY	740	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	759
Db	2237	TTTACCCGAGTAGGAGTCTGGAACAAACACAGACAGATAGTAACATGTTTCTGTATTG	2296
QY	760	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	779
Db	2297	TTTCTAGTCCTTGAATATTATTTCTTTTATACATTTTCTCCAGGCGTTCACATTTGGC	2356
QY	780	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	799
Db	2357	AAAGCTGGGAGATCCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTGAGA	2416
QY	800	GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	819
Db	2417	CAGGATGTCAGCTGGTGTGATGACCTAAAAACACCACCTGGAGCATTTGACAAACGCGTT	2476
QY	820	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	839
Db	2477	GCCAAATGATCGGCTCAAGTTAAAGGGGCTATAGTGTCCAGCGTGTGTCATTAACCCAG	2536
QY	840	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr	859
Db	2537	AAATACCAATCTTGGCAGGCATATTATATCTTAATCTATGTTGGCAATTAACA	2596
QY	860	LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet	879
Db	2597	CTTTTACTCTTACCAATTGACCATCATGTCAATAGCAGAGTGTGTGAATGAAATG	2656
QY	880	LeuSerGlyGlnAlaLeuLysAspLysGlyGluLeuGlyAlaGlyLysIleAlaThr	899
Db	2657	TTGTCTGGACACACTGAAGATAGAAGAGCTAGAAGAGCTGGGAAGATTGGTACA	2716
QY	900	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr	919
Db	2717	GAAGCCATCGAAACCTTCGAACTGTCTTCTTGTGCTGGGAGCAGAGTTTGAATC	2776
QY	920	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	939
Db	2777	ATGTATGCACAGATTTCAGTACCATACAGAAACCTTTTGAGGAAGACACATCTTC	2836
QY	940	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	959
Db	2837	GGGGTCTCATTTTCTATCACCCAGGCATGATGATATTTTTCCTATGCTGGCTGTTCCGG	2896
QY	960	PheGlyAlaTyrIleuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe	979
Db	2897	TTTGGTGCCTACTTGGTGGCAATGAGTTCATGAACTTTCAGAGATGCTCTTTTGGTATT	2956
QY	980	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	999
Db	2957	TCAGCTATTGTCTTTGGTGCCATGGCAGTGGGCGAGTCACTTCTTCTCTGACTACT	3016
QY	1000	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1019

Db 3017 GCCAAAGCCAAAGTATCAGCACCCACGTCATCATGATTCATGAAAAAGCCCTCTGATT 3076
QY 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
Db 3077 GACAGCTACAGCCCTCAGGCCCTCAAGCCAAATACAGTTGGAAGGAAATGTGACATTTAAT 3136
QY 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
Db 3137 GAGGTGCTGTTCAACTATCCCACTCGACACAGACATCCCGTGTCTCCAGGGGCTGAGCCCTC 3196
QY 1060 GluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
Db 3197 GAGGTGAAGAAGGCGCAGACGCTGGCCCTCGTAGGTAGCAGTGGCTGGGAAGAGCACA 3256
QY 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuLeuAspGly 1099
Db 3257 GTTGTTTCAGCTCTAGAGCGCTTCTATGACCCCTTGGCTGGTTCAGTGTCTAATTGATGCC 3316
QY 1100 LysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSerGln 1119
Db 3317 AAAGAGATAAGCACCTGAATGTCCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAG 3376
QY 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
Db 3377 GAGCCCATCTGTTGACTGCAGCATTCGCGAACAATTCCTATGGAGACAACAGCCGG 3436
QY 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1159
Db 3437 GTCGTATCATGAAGAGATTATGAGGCGAGCCCAAGGAGGCCAACATACACCACCTTCATC 3496
QY 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
Db 3497 GAGACATCTCCTGAGAAATACACACAGCAGTAGAGACAAGAACCCAGCTCTCTGCT 3556
QY 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
Db 3557 GGCCAGAAACAGCGCATTCGCATAGCTCGCGCTCTGTGTAGACAGCGCTCATATTTGCTT 3616
QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
Db 3617 TTGATGAAGCTACATCAGCTCTGGATACAGAAGTGAAAAGTTGTCCAGAAGCCCTG 3676
QY 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 3677 GACAAAGCCAGAGAAGCGCCACCTGCTGATGTGATCGCCACCGCTGTCCACCATCCAG 3736
QY 1240 AsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
Db 3737 AATGCAGATTTAATAGTGGTGTTCAGAAATGGCAAGTCAAGGAGCAGTGGCACACATCAA 3796
QY 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 3797 CAGCTGTGGCCCAAGAAAGCATCTATTTTCCATGTGTGATGCTCCAGGCTGGAGCAAG 3856
QY 1280 Arg 1280
Db 3857 CGC 3859
RESULT 3
AAD03505 standard; cDNA; 4279 BP.
XX AC AAD03505;
XX DT 13-JUN-2001 (first entry)
XX DE Dog P-glycoprotein (PGP) allelic variant (Genotype B) cDNA.
XX KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
KW MDR1; drug bioavailability; transgenic animal; genetic model; ss.
XX OS Canis familiaris.

XX Key Location/Qualifiers
PH 17..3862
FT /tag= a "dog P-glycoprotein (PGP) allelic variant
FT /product= (Genotype B) protein"
FT replace (91, T)
FT allele /tag= b
FT /tag= replace (607, C)
FT allele /tag= C
XX WO200123540-A2.
XX PN 05-APR-2001.
XX PD 28-SEP-2000; 2000WO-US26767.
XX PF 28-SEP-1999; 99US-0156510.
XX PR (GENT-) GENTEST CORP.
XX PA Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX PI WPI: 2001-235373/24.
XX DR P-PSDB; AAE00309.
XX DR New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX PT for determining the bioavailability of drugs and for screening for dog
XX PT PGP inhibitors -
XX PS Claim 9; Page 93-99; 111pp; English.
XX CC The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDR1) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) allelic variant
CC (Genotype B) cDNA. The PGP enzyme functions as an efflux pump
CC exporting small molecules across the cell membrane. This enzyme
CC is a member of the ABC transporter family.
XX Sequence 4279 BP; 1296 A; 833 C; 1008 G; 1142 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0 Length: 4279
Score: 6425.50 Matches: 1273
Percent Similarity: 99.69% Conservative: 4
Best Local Similarity: 99.38% Mismatches: 3
Query Match: 99.36% Indels: 1
DB: 22 Gaps: 1
US-09-672-725C-4 (1-1280) x AAD03505 (1-4279)
QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 17 ATGGATCCTGAAGGAGCGCTTAAGGGGAGTGCAGAGAGAACTTCTGGAAAAATGGCAAA 76
QY 21 LysSer---LysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
Db 77 AAAAGTAAAAAAG 136
QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 59
Db 137 CGTATTCAATTGGCTTGATAGTTGTATATGTTGGTGGGACAATGGCTGCCATCATC 196

QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
| | | | |
Db 197 CATGGAGTGCACCTCCCTCATGATGCTGGTTTTGGAAACATGACAGATAGCTTGGCA 256
QY 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
| | | | |
Db 257 AATGAGGAATTTCAAGAAACAAACATTTTCCAGTTATAATTAATGAAGATTAACGAAC 316
QY 100 AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyrTyr 119
| | | | |
Db 317 AATACACAACTTTCATCCACCTCTGGAGGAGGAATGACCAGTATGCTTATTATTAC 376
QY 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheThrCysLeu 139
| | | | |
Db 377 AGTGGATCGGTGCTGGCGTCTGGTGGCTTTACATCCAGGTTCATTTCTGGTCCCTG 436
QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
| | | | |
Db 437 GCAGCAGGAAGACAGATACTCAAAATTAGAAAAACAATTTTTCATGCTATCATCGCACAG 496
QY 160 GluIleGlyTyrPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
| | | | |
Db 497 GAGATTGGCTGGTTGACGTGCATGACGTTGGGGAGCTTAACCCGGCTTCACAGACGAT 556
QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
| | | | |
Db 557 GTCTCCAAAATCAATGAAGGAATTTGGCGACAAAATTTGGAATGTCTTTCAATCAATAGCA 616
QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTyrLysLeuThrLeuValIle 219
| | | | |
Db 617 ACATTTTTCACCGGTTTTATAGTGGGGTTTTACACGTGGTGGGAAGCTAAACCCCTTGTGATT 676
QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTyrAlaLysIleLeuSerSer 239
| | | | |
Db 677 TTGGCCATCAGCCCTGTTCTTGGACTTCAGCGCCCATCTGGCAAGATATATCTTCA 736
QY 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValaGluGluValLeu 259
| | | | |
Db 737 TTTACTGATAAAGAACTTTGGCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAGTCTTA 796
QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
| | | | |
Db 797 GCAGCAATCAGAACTGTGATGGCTTTGGAGGACAAAGAAAGAACTTGAAGGTACAAAC 856
QY 280 LysAsnLeuGluAlaLysArgIleGlyIleLysLysLysAlaIleThrAlaAsnIleSer 299
| | | | |
Db 857 AAAAAATTTAGAAGAACTAAAGGAATTTGGATAAAGAAAGCTATCACGGCCCAACATTTCT 916
QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTyrPtyrGlyThr 319
| | | | |
Db 917 ATTGGTGCCGCTTTCATTATGTATGATCATGATCATATGCTTGGCTTTCTGGTATGGGACC 976
QY 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
| | | | |
Db 977 TCCTTGGTCTCTCCAGTGAATATCTATGGACAAGTACTCACCTGTCTTCTTCTGTGA 1036
QY 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
| | | | |
Db 1037 TTAATTTGGGGCTTTTAGTATTGGACAGGCATCCCCAAGCAATTTGAAGCATTTGCAACGCCA 1096
QY 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
| | | | |
Db 1097 AGNAGGACAGCTTATGAATCTTCAGGATAATTTGACATAAACCACCAAGCATTTGACAGCAT 1156
QY 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
| | | | |
Db 1157 TCGAAGAGTGGACATAAACACAGATAATTAAGGGAAATTTGGAAATTCAAAATATGTTTAC 1216
QY 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
| | | | |
Db 1217 TTCAGTTACCTTCTCGAAAAAGAGTTAAGATCTTAAGGGTCTCAACCTGAAGGTTTAC 1276

QY 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
| | | | |
Db 1277 AGTGGCCAGACAGTGGCGCTGTTGGGAACAGTAGTGGCTGCGGGAAGACACACCGTGCAG 1336
QY 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
| | | | |
Db 1337 CTGATGCAGAGGCTCTATGACCCCCACAGATGGCATGTGCTGTATTTATGTAGACAGGACATT 1396
QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
| | | | |
Db 1397 AGGACATAAATTAAGGCATCTTCGGAAATTACTGGTGGTGGTGGTGGTGGTGGTGGTGG 1456
QY 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
| | | | |
Db 1457 TTGTTTGGCCACCCAGATAGCTGAAACATTCGCTATGGCCGCGAAATGTACCATGGAT 1516
QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
| | | | |
Db 1517 GAGATTGAGAAAGCTGTTAAGGAAGCCAAATGCCATGATTTTATCATGAACACTACCTAAT 1576
QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
| | | | |
Db 1577 AAATTTGACACTCTGTTGGAGAGAGAGGGGCCAGCTGAGTGGTGGCAGAAACACAGAGA 1636
QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
| | | | |
Db 1637 ATCCCAATGCTCGGGCCCTGGTTCGCAACCCCAAGATTTCTTCTGCTGGTGGCAAGG 1696
QY 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
| | | | |
Db 1697 TCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGTGGCCCTGGATGAAGGCCAGAAA 1756
QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
| | | | |
Db 1757 GGGCGGACTACCAATGATGATAGCTCATCGTTTGTCTACAGTTCGTAATGCCATGCTCAT 1816
QY 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
| | | | |
Db 1817 GCTGGTTTTCATGATGGATCATGTGGAGAAAGGAATCATGATGAATCATGAAGAGAG 1876
QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
| | | | |
Db 1877 AAGGGCATTTTACTCAAACTTGTCACAAATGCAGACAAAGAGAAATGAATTCAGTTAGAA 1936
QY 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
| | | | |
Db 1937 AATGCCACTGGTGAATCCAAAAGAGGACTGAATGAGAATGTACCTCCAGTTTCTTCTTCTGG 1996
QY 660 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
| | | | |
Db 1997 GGGTCCAGTTTAATAAAGAGAGATCACTCCAGAGATATACATGCACCAAGGCCAA 2056
QY 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTyr 699
| | | | |
Db 2057 GACAGAAAGCTTGGTACAAAAGAGGACTGAATGAGAATGTACCTCCAGTTTCTTCTTCTGG 2116
QY 700 ArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 719
| | | | |
Db 2117 AGGATTTCTGAAGCTGAACCTCAACTGAATGGCCCTTATTTGGTGGTATATTTTGTGCT 2176
QY 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
| | | | |
Db 2177 ATTATAACGGAGGCTGCAACCCAGCAATTTCAATAATATTTTCAAGGATTTATAGGATC 2236
QY 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
| | | | |
Db 2237 TTTACCCGAGATGAGGATCTCTGAAACAAAACAGACAGAAATAGTAACATGTTTCTGTATTG 2296
QY 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
| | | | |
Db 2297 TTTCTAGTCTTGGAAATATTTCTTTTATACATTTTCTCCAGGGCTTCACATTTGGC 2356
QY 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799

|||||
Db 2357 AAAGCTGGGGAGATCCTACTAAGCGGCTTCGATACATGGTTTTTCAGATCCATGCTGAGA 2416
Qy 800 GlnAspValSerTirPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
Db 2417 CAGATGTCAGCTGGTTGATGACCTTAAACACACACTGGAGCATTGACAAACCGGCTT 2476
Qy 820 AlaAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
Db 2477 GCCAATGATCGCGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCTGCTCATACCCAG 2536
Qy 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleIleGlyTirPInLeuThr 859
Db 2537 AATATAGCAAACTCTGGGACAGCATTTATATATCTTAATCTATGGTTGGCAATTAACA 2596
Qy 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
Db 2597 CTTTTACTCTTAGCAATGTACCCATCATTCGAATAGCAGGAGTTGTTGAAATGAAATG 2656
Qy 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
Db 2657 TTGTCTGGACACACACTGAAAGATAGAAGAGCTAGAAGGAGCTGGAGATGCTACA 2716
Qy 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
Db 2717 GAAGCCATCGAAAACTTCCGAACCTGTTGTTCTTGACTCGGGAGCAGAGTTTGAATAC 2776
Qy 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
Db 2777 ATGTATGCACAGAGTTTGCAAGTAGTACCATACAGAAACTCTTTGAGGAAGCACACATCTTC 2836
Qy 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
Db 2837 GGGGTCTCATTTTCTATCACCCAGGCAATGATGATTTTTCCTATGCTGGCTGTTCCCG 2896
Qy 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 979
Db 2897 TTTGGTGCTACTTGTGGCAATAGTTCATGAACCTTCAGGATGCTCTTTGGTATTC 2956
Qy 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
Db 2957 TCAGCTATTTGCTTTGGTGCCATGGCAGTGGGAGGTGATGTTTCTGCTGCTGACTAT 3016
Qy 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleLeuLysSerProLeuIle 1019
Db 3017 GCCAAGCCAAAGTATCAGCAGCCAGCTCATCATGATCATGTAATAAAGGCGCTCTGATT 3076
Qy 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
Db 3077 GACAGCTACAGCCCTCAGCGGCTCAAGCCAAATACGTTGGGAAGAAATGTGACATTTAAT 3136
Qy 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
Db 3137 GAGGTCGTGTCACTATCCCACTCGACACACATCCCGTGCTCCAGGGGCTGAGCCTC 3196
Qy 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
Db 3197 GAGGTGAAGAGGGCCAGACGCTGGCCCTCGTAGTAGCAGTGGCTGTGGGAAGAGCACA 3256
Qy 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
Db 3257 GTTGTTCAGCTCCTAGAGCGCTTCTATGACCCCTTGGCTGGTTGAGTCAATTTGATGGC 3316
Qy 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
Db 3317 AARGAGATAAGCACTGAATGTCAGTGCTCCAGTGCTCCGAGCACACCTGGGCATCGTCTCAG 3376
Qy 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
Db 3377 GAGCCCATCTCTTTGACTGACGATTCGCGAAGACATGCGTATGGAGACAACAGCCGG 3436
Qy 1140 ValValSerHisGluLeuIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1159
|||||

Db 3437 GTCGTATCACATGAAGAGATTATGCAGCGCAAGAGGCGCAACATACACCACTTCATC 3496
Qy 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
Db 3497 GAGACACTCCCTGAGAATAACACACACAGAGTAGGACAAAGAACCCAGCTCTCTGGT 3556
Qy 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
Db 3557 GGCCAGAAACAGCGCATTCGCATAGCTCGGCTCTTGTAGACAGCCTCATATTTGCTT 3616
Qy 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
Db 3617 TTGGATGAAGTACATCAGCTCTGGATACAGAAAGTGAAGAGTTGTCGAAGAGCCCTG 3676
Qy 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 3677 GACAAGCCAGAGAAGCGCCACCTGCATTGTATCGCCACCGCTGTCCACCATCCAG 3736
Qy 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
Db 3737 AATGCAGATTTAATAGTGTGTTTCAGAAATGCAAGGAGCATGGCCACACATCAA 3796
Qy 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 3797 CAGCTGCTGGCCAGAAAGGCATCTATTTTCCATGGTCAGTGTCCAGGCTGGAGCAAAG 3856
Qy 1280 Arg 1280
|||
Db 3857 CGC 3859
RESULT 4
AAD03506
ID AAD03506 standard; cDNA; 4279 BP.
XX
AC AAD03506;
XX
DT 13-JUN-2001 (first entry)
XX
DE Dog P-glycoprotein (PGP) allelic variant (Genotype D) cDNA.
KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
KW MDRI; drug bioavailability; transgenic animal; genetic model; ss.
XX
OS Canis familiaris.
XX
FH Key Location/Qualifiers
FT CDS 17..3862
FT /product= "Dog P-glycoprotein (PGP) allelic variant
FT (Genotype D) protein"
FT replace (91, T)
FT /tag= b
FT replace (607, C)
FT /tag= c
FT replace (1001, T)
FT /tag= c
FT replace (3458, A)
FT /tag= c
XX
PN W0200123540-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26767.
XX
PR 28-SEP-1999; 99US-0156510.
XX
PA (GENT-) GENTEST CORP.
XX
PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX WPI; 2001-235373/24.
DR P-PSDB; AAE00310.
DR

XX	New dog p-glycoproteins (PGP) and their encoding nucleic acids, useful
PT	for determining the bioavailability of drugs and for screening for dog
PT	pgp inhibitors -
PT	

PS Claim 9; Page 102-107; 111pp; English.

The invention relates to dog P-glycoprotein (PGP) also referred as multidrug transporter (MDR1) and nucleic acids encoding them. The invention also includes fragments and biologically functional variants of dog P-glycoprotein. PGP and their nucleic acids are useful for determining the bioavailability of drugs and for screening PGP inhibitors. They are useful for the diagnosis and treatment of conditions characterised by PGP activity, by reducing or increasing PGP activity in a cell. PGP nucleic acids are used as oligonucleotide probes. Complements of PGP nucleic acids are useful as antisense oligonucleotides, to induce a PGP 'knockout' phenotype. They are used to prepare a non-human transgenic animal, which are valuable as genetic models for human diseases.

CC The present sequence is dog P-glycoprotein (Pgp) allelic variant
CC (Genotype B) cDNA. The Pgp enzyme functions as an efflux pump
CC exporting small molecules across the cell membrane. This enzyme
CC is a member of the ABC transporter family.

SQ Sequence 4279 BP; 1296 A; 833 C; 1009 G; 1141 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	4279
Score:	6425.50	Matches:	1273
Percent Similarity:	99.69%	Conservative:	4
Best Local Similarity:	99.38%	Mismatches:	3
Query Match:	99.36%	Indels:	1
DB:	22	Gaps:	1

US-09-672-725C-4 (1-1280) x AAD03506 (1-4279)

QY	1	MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPhetrrPlysMetGlyLys	20
DB	17	ATGGATCTCTGAAGGAGGCCCTAAGGGGAGTGCAGAGAAGAACTCTCGAAANTGGGCAA	76
QY	21	LysSer---LysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe	39
DB	77	AAAGCTAAAAAAGAAGAGAAGAAAAACCAACCTGTCAGCACGTTTGTCAATGTTT	136
QY	40	ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle	59
DB	137	CSCTATTCAANTGGCTTGATAGTGTATATGTTGGTGGGCAATGGCTGCCATCATC	196
QY	60	HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	79
DB	197	CATGGAGCTGCACCTCCCTCATGATGCTGGTTTTTGGAAACATGACAGATAGCTTTGCA	256
QY	80	AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	99
DB	257	AATGCAGGAATTTCAAGAAACAAACCTTTTCCAGTTATATTAATGAAGTATTACGAAC	316
QY	100	AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr	119
DB	317	AATACACACAACTTCATCACCCATCTGGAGGAGAAATGACACAGTATGCCTATTATTAC	376
QY	120	SerGlyIleGlyAlaGlyValLeuValAlaAlaTatyrIleGlnValSerPheTrpCysLeu	139
DB	377	AGTGGGATCGTGCTGGCGTGTGGTGCCTTACATCCAGGTTTCATTCTGGTCCCTG	436
QY	140	AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln	159
DB	437	GCACGAGGAAGACAGATACTCAAAATTAGAAAAAATTTTTTCATGCTATCATGCGACAC	496
QY	160	GlnIleGlyTrpPheAspValHisAspValGlyGlnLeuAsnThrArgLeuThrAspAsp	179
DB	497	GAGATTGGCTGGTTTGCATGTCATGACGTTGGGAGCTTAACACCGGCTCACACAGAT	556

1637 ATCCGATTGCTCGGCCCTGGTTCGAACCCAGAGATCTTCTGCTGGATGAGCAACG 1696
 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
 1697 TCAGCTCTGGACACTGAAGTGAAGCAGTGGTTCAGGTGGCCCTGGATAAGGCCAGAAAA 1756
 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
 1757 GGCCGGACTACCATTTGTGATAGCTACCTGTTGTCTACAGTTTCGTAATCCGATGTCATT 1816
 600 AlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
 1817 GC7GGTTTTGATGATGGAGTCATTGTGGAGAAAGGAATCATGATCACTGAAGAG 1876
 620 LysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
 1877 AAGGCATTTTACTTCAAACTTGTCAAAATGCAGAAAGAGGAATGAAATTCAGTTAGAA 1936
 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
 1937 AATGCCACTCGTGAATCCAAAGTGAAGAGTATGCTTGGAAATGTCTCCAAAGATTCA 1996
 660 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
 1997 GGGTCCAGTTTATATAAAGAGATCAACTCGCAGGAGTATACATGCACCACAAGCCAA 2056
 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrp 699
 2057 GACAGAAAGCTTGGTACAAAAGAGGACTTGAATGAGAATATACCTCAGCTTCCCTCTGG 2116
 700 ArgIleLeuLysLeuAsnSerThrGluTrpProThrPheValValGlyIlePheCysAla 719
 2117 AGGATTCTGAAGCTGAACTCACTGAATGGCCCTATTTTGTGGTATATTTTGTGCT 2176
 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
 2177 ATTATAAGCGAGGCCCTGCAACCAAGCATTTTCAATAATATATTTCAAGGATATAGGGATC 2236
 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
 2237 TTTACCCGAGATGAGGATCTCTGAACAAAAGCAGCAAGATAGTAACTGTTTCTGTATTG 2296
 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
 2297 TTTCTAGTCCCTGGANATATTTCTTTATTTACATTTTCTCCAGGCTTCACATTTGGC 2356
 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
 2357 AAAGCTGGGAGATCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTGAGA 2416
 800 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
 2417 CAGGATGTCAGCTGGTTGTATGATCCCTTAAACCAACCACTGGAGCATTTGACACCAGCCT 2476
 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
 2477 GCCATGATCGGCTCAAGTTAAAGGGCTATAGTTCACAGCTTGTGTCATTACCCAG 2536
 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 859
 2537 AATATAGCAAAATCTTTGGACAGGCATATATATCTTAATCTATGGTGTGGCAATTAACA 2596
 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
 2597 CTTTACTCTTAGCAATTTGTACCCATCATTCGAATAGCAGGAGTTGTTGAAATGAAATG 2656
 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
 2657 TTGTCTGGACAAGCACTGAAGATAGAAGAGCTAGAAAGAGCTGGGAAGATTGCTACA 2716
 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919

2717 GAAGCCATCGAAAACTTCCGAACTGTTGTTCTTTGACTCGGGAGCAGAAAGTTTGAATAC 2776
 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
 2777 ATGTATGCACAGAGTTTGCAGTACCATACAGAACTCTTTGAGGAAAGCACACATCTTC 2836
 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
 2837 GGGGTCTCATTTTCTATCACCAGCAATGATGATTTTCTCTATGCTGCTGTTCCGG 2896
 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 979
 2897 TTTGGTGCTACTTTGGTGGCAATGAGTTCAAGACTTCAGGATGTTCTTTTGGTATTC 2956
 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
 2957 TCAGCTATTGCTTTGGTGCCATGSCAGTGGGCGAGGTTCAGTTTCATTTGCTCCTGACTAT 3016
 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
 3017 GCCAAAGCCAAAGTATATCAGCAGCCCATCATCATGATCATTTGAAAAAAGCCCTCTGATT 3076
 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
 3077 GACAGCTACAGCCCTCACGGCTCAAGCCAAATAGCTTGGAAAGGAATGTGACATTTAAT 3136
 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
 3137 GAGGTCGTGTTCAACTATCCACTCGACAGACATCCCGTCTCCAGGGGCTGAGCCTC 3196
 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
 3197 GAGGTGAAGAGGGCCAGACGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGACACA 3256
 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
 3257 GTTGTCAGCTCCTAGAGCGCTTCTATGACCCCTTGGTGGTTCAGTGCATTTGATGGC 3316
 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
 3317 AAAGAGATAAAGCACCTGAATGTCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAG 3376
 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
 3377 GAGCCCATCCTGTTGACTGACGATTGCCGAGAACTTGCCTATGGAGACAACACCGCG 3436
 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1159
 3437 GTCGTATCATGAAGAGATTGTGAGGCGCAAGGAGGCCAACATACACACTTCATC 3496
 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
 3497 GAGACACTCCCTGAGAAATACAAACACGAGTAGGAGACAAGAACCCAGCTCTCTGCT 3556
 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
 3557 GGCCAGAAACAGCGCATTCGCATAGCTCGCGCTCTTGTGTAGACAGCCTCATATTTGCTT 3616
 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
 3617 TTGGATGAAGCTACATCAGCTCTGGATACAGAAAGTGAAGAGGTTGTCAGAAGGCCCTG 3676
 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
 3677 GACAAGCCAGAGAGGCCGCCACCTGCATTTGATCGCCACCGCTTGTCCACCATCCAG 3736
 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
 3737 AATGCAGATTAATAGTGGTGTTCAGATGTCAGAAAGTCAAGGAGCATGGCACACATCAA 3796
 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
 3797 CAGTGTCTGCCCAAGAAAGGCATCTATTTTCCATGGTCAAGTGTCCAGGCTGGAGCAAG 3856

QY 1280 Arg 1280
 Db 3857 CGC 3859
 RESULT 5
 AAD03488
 ID AAD03488 standard; cDNA; 4279 BP.
 XX
 AC AAD03488;
 DT 13-JUN-2001 (first entry)
 XX
 DE Dog P-glycoprotein (PGP) cDNA #1.
 XX
 KW Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
 drug bioavailability; transgenic animal; genetic model; ss.
 OS Canis familiaris.
 XX
 FH Key Location/Qualifiers
 FT CDS 17..3862
 FT /*tag= a
 FT /product= "Dog P-glycoprotein (PGP) #1"
 XX
 PN WO200123540-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26767.
 XX
 PR 28-SEP-1999; 99US-0156510.
 XX
 PA (GENT-) GENTEST CORP.
 XX
 PI Stocker PJ, Steelmel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 XX
 DR WPL; 2001-235373/24.
 DR P-PSDB; AAE00303.
 XX
 PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 for determining the bioavailability of drugs and for screening for dog
 PGP inhibitors -
 XX
 PS Claim 3; Page 58-63; 11pp; English.
 CC The invention relates to dog P-glycoprotein (PGP) also referred
 as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (PGP) cDNA. This
 CC sequence is also referred as Genotype C cDNA. The
 CC PGP enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.
 XX
 SQ Sequence 4279 BP; 1294 A; 834 C; 1008 G; 1143 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 4279
 Score: 6420.50 Matches: 1272
 Percent Similarity: 99.61% Conservative: 4
 Best Local Similarity: 99.30% Mismatches: 4
 Query Match: 99.28% Indels: 1

DB: 22 Gaps: 1
 US-09-672-725C-4 (1-1280) x AAD03488 (1-4279)
 QY 1 MetAspProGluGlyGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
 Db 17 ATGGATCTCTGAAGAGGCCGTAAAGGGAGTGCAGAGAAAGAACTCTGGAAATGGCAAA 76
 QY 21 LysSerLysLys---GluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
 Db 77 AAAAGTAAATAAATGAGAGAAAGAAACCAACTCTGCACGAGTTTGAATGTTT 136
 QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 59
 Db 137 CGCTATTCAAATTTGGCTTGATAGTTGATATGTTGGTGGGACAATGGCTGCATCATC 196
 QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
 Db 197 CATGAGCTGCACCTCCCTCTCATGATGCTGGTTTTTGGAAACATGACATAGCTTTGCA 256
 QY 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
 Db 257 AATCAGAGAAATTTCAAGAAACAAAACCTTTTCCAGTTTATAATTAATGAAAGTATTACGAC 316
 QY 100 AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr 119
 Db 317 AATACACACATTTCATCAACCATCTGGAGGAGAAATGACCACGCTATGCCTATTATTAC 376
 QY 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGluValSerPheTrpCysLeu 139
 Db 377 AGTGGGATCGTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 436
 QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
 Db 437 GCAGCAGGAGACAGATACTCAAAATTGAAACAAATTTTTTCATGCTATCATGCGACAG 496
 QY 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
 Db 497 GAGATTGGCTGGTTTACGTGTCATGACGTGGGAGCTTAACACCCGCTCACACAGCAT 556
 QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
 Db 557 GTCTCAAAATCAATGAAGSAATTGGCGACAAAATTGGATGTTCTTTCACTCAATAGCA 616
 QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
 Db 617 ACATTTTTTCCCGGTTTTTATAGTGGGGTTTACACGTGGTTGGAAGCTAACCCTTGTGAT 676
 QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
 Db 677 TTGGCCATCAGCCCTGTTCTTGACCTTTTCAGCCCATCTGGGCAAGATACTACTCTCA 736
 QY 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
 Db 737 TTTACTGTATAAAGAACTCTTGGCCCTATGCAAAAGCTGGAGCAGTAGCTGAAGAAAGTCTTA 796
 QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
 Db 797 GCAGCAATCAGAACCTGATGTCCTTTGGAGGACAAAAGAAAGAACTTGAAGAGTACAAC 856
 QY 280 LysAsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
 Db 857 AAAAATTTAGAAGAGCTAAAGGAATTTGGGATAAAGAAAGCTATCAGCGGCACATTTCT 916
 QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
 Db 917 ATTGTGCGCTTCTTATTGATCTATGATCATATGCTGCTGGCTTCTGCTGATGGGACC 976
 QY 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
 Db 977 TCCTTGGTCCCTCTCAGTGAATATTTCTATGGACAAGTACTACTGCTCTCTTTCTGTA 1036
 QY 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359

1037 TTAATTTGGGGCTTTTATGATTTGGACAGGCATCCCAAGCATTTGAAGCATTTGCCAACCACA 1096
 360 AtcGlyAlaAtyrGluIlePheLysIleLeuAspAsnLysProSerIleAspSerTyr 379
 1097 AGAGGACAGCTTATGAAATCTTCAGATAATGACATAAACCAAGCATTTGACAGCTAT 1156
 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
 1157 TCGAAGAGTGGACATAAACACAGATAATATAAGGGAAATTTGGAATTTCAAAAATGTTCCAC 1216
 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
 1217 TTCAGTTACCTTCTCGAAAAGAAAGTTAAGATCTTAAAGGGTCTCAACCTGAAGGTTGAG 1276
 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
 1277 AGTGGGAGACAGTGGGCTGGTTGGACAGTGGCTGGCGGAAGACGACGCGGTGCAG 1336
 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
 1337 CTGATGCAGAGGCTATGATGCCACAGATGGCATGGTCTGTATTGATGGACAGGACATT 1396
 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
 1397 AGGACCAATAATGTAAGGCATCTTCGGGAAATTTACTGGTGTGGTGTGAGTCCAGGAGCTG 1456
 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
 1457 TTGTTGCCACACAGATAGCTGAAACATTCGCTATGGCCGCAAAATGTCACCATGGAT 1516
 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
 1517 GAGATTGAGAAAGCTGTTAAGGAAGCCAAATGCCCTATGATTTATCATGAACCTACCTAAT 1576
 520 LysPheAspThrLeuValGlyLysArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
 1577 AAATTTGNACACTCTGGTTGGAGAGAGGGGCCAGCTGAGTGGTGACAGAAACAGAGA 1636
 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuAspGluAlaThr 559
 1637 ATCGCCATTGCTCGGGCCCTGGTTGCAACCCCAAGATTTCTGCTGGATGAGGCAACG 1696
 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
 1697 TCAGCTCTGGACACTGAAGTGAAGCAGCTGTTTCAGGTGGCCCTGGATAAGGCCAGAAAA 1756
 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
 1757 GSCCGGACTACCATTTGATAGCTCATCGTTTGTCTACAGTTCGTAATGCGGATGTCATT 1816
 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
 1817 GCTGCTTTTGATGATGGAGTCATTGTGGAGAAGAAATCATGATGAATCATCAAGAGAG 1876
 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
 1877 AAGGCCATTACTTCAAACTTGTCAATGCAGACAGCAAGGAAATGAAATTTGAGTTGAA 1936
 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
 1937 AATGCCACTGGTGAATCCAAAAGTGAAGTGGTTCGTTGAAATGTCCTCCAAAAGATTCA 1996
 660 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
 1997 GGTCCAGTTTAAATAAAGAACATCACTCCGAGGATATACATGCACCAAGGCCAA 2056
 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrp 699
 2057 GACAGAAAGCTTGGTACAAAGAGGACTTGAATGAGAATGATACCTCCAGTTCTCTTCGG 2116
 700 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719

2117 AGGATTCGAAGCTGAACCTCAACTGAATGGCCTTATTTTGTGGTGGTATATTTTGTGCT 2176
 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
 2177 ATTATAAACGGAGCCCTGCAACACAGCATTTTCAATAATATTTTCAAGGATTATAGGATC 2236
 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
 2237 TTTACCCGAGATGAGGATCCTGAAACAAACGACAGAAATAGTAACATGTTTCTGTATTG 2296
 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
 2297 TTTCTAGTCCTTGGAAATTTATTTTATTACATTTTCTCCAGGGCTTCACATTTGGC 2356
 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
 2357 AAAGCTGGGAGATCTTCACCTAAGCGGCTTCGATACATGTTTTCAGATCCATCCTGAGA 2416
 800 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
 2417 CAGGATGTCAGCTGTTTGTATGACCCCTAAACACACACTGGAGCATTCACAAACAGGCTT 2476
 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
 2477 GCCAATGATGGGCTCAAGTTAAAGGGCTATAGTTCAGGCTTGTCTCATACCCAG 2536
 840 AsnIleAlaAsnLeuGlyThrGlyIleIleLeuSerLeuIleTyrGlyTrpGlnLeuThr 859
 2537 AATATAGCAAAATCTTGGACAGGCATTTATATCTTAACTATGTTGGCAATTAACA 2596
 860 LeuLeuLeuLeuAlaIleValIleProIleAlaIleAlaGlyValValGluMetLysMet 879
 2597 CTTTACTCTTACCAATTTGATCCCATCATTTGAAGGGCTATAGTTCAGGCTTGTGAATGAAATG 2656
 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
 2657 TTGCTGGACAGACTGAAAGTAAAGAAAGAGCTAGAAAGAGCTGGGAAGATTTCTACA 2716
 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
 2717 GAAGCCATCGAAACCTCCGAACTGTTTCTTTGACTCGGGAGCAGCAAGTTTGAATAC 2776
 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
 2777 ATGATGACAGAGTTTGCAGTACCATACAGAACTCTTTGAGAAAGACACATCTTC 2836
 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
 2837 GGGGTCTCATTTTCTATCACCCAGCAATGATGATTTTTCCTATGCTGGCTGTTTCCGG 2896
 960 PheGlyAlaTyrIleuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 979
 2897 TTTGGTGCCTACTTGGTGCAATGAGTTCATGAACCTTTTCAGGATGTTCTTTGGTATTC 2956
 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
 2957 TCAGCTATTGCTTTTGGTGCATGGCAGTGGGAGGCTCAGTTTTCATTTGCTCGTACTAT 3016
 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
 3017 GCCAAAGCCAAAGTATCAGCAGCCCGCATCATCATCATTTGAAAAAAGCCCTCTGATT 3076
 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
 3077 GACAGCTACAGCCCTCAGGCCCTCAAGCCAAATACGTTGGAAGAAATGTGACATTTAAT 3136
 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
 3137 GAGGTCTGTTTCAACTATCCACTTCGACACAGACATCCCGTGTCTCCAGGGGCTGAGCCCT 3196
 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
 3197 GAGGTGAAGAGGGCCAGACCTGCTGGCCCTCGTAGGTAGCAGTGGCTGTGGGAAGAGCACA 3256

```
QY 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
|||||
Db 3257 GTTGTTCAGCTCTAGACCGCTTCTATGACCCCTTGCTGCTCAGTGCCTAAATGATGGC 3316

QY 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
|||||
Db 3317 AAAGAGATAAAGCACCTGAATGTCCAGTGGCTCCGAGCACACCTGGGCATCGTGTCTCAG 3376

QY 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
|||||
Db 3377 GAGCCCATCTCTGTGTGCTGAGCATTTGCCGAGAACAATTCCTATGGAGACAACAGCCGG 3436

QY 1140 ValValSerHisGluGluIleAlaIleAlaAlaLysGluAlaAsnIleHisHisPheIle 1159
|||||
Db 3437 GTCGTATCACATGAAGATATTGCGAGCGAGCCAGGAGGCCCAATACACCACTTCATC 3496

QY 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
|||||
Db 3497 GAGACACTCCCTGAGAATAACACACAGAGTAGGAGACAAGGAACCCAGCTCTCTGCT 3556

QY 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
|||||
Db 3557 GCCAGAAACAGCCATTCGATAGCTCGGCTCTTGTGTAGACAGCTCATATTTGCTT 3616

QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
|||||
Db 3617 TTGGATGAAGCTACATCAGCTCGGATACAGAAGTGAAGAGGTGTCTCCAGAACGCCCG 3676

QY 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
|||||
Db 3677 GACAAAGCCAGAGAAAGCCGACCTGCTATGTGATCGCCAGCGCTGTCCACCATCCAG 3736

QY 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValIysGluHisGlyThrHisGln 1259
|||||
Db 3737 AATGCAGATTAAATAGTGGTTCAGATGGCAAGTCAAGAGCATGGCCACATCAA 3796

QY 1260 GlnLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
|||||
Db 3797 CAGCTGCTGCCCCAGAAAGGACATCTATTTTCCATGGTCAGTGTCCAGGCTGGAGCAAG 3856

QY 1280 Arg 1280
|||
Db 3857 CGC 3859

RESULT 6
AAZ49332
ID AAZ49332 standard; cDNA; 3860 BP.
XX
AC AAZ49332;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human wild-type multidrug resistance-1 (MDR-1) cDNA.
XX
KW Multidrug resistance; MDR-1; P-glycoprotein;
KW transmembrane efflux pump; haematopoietic stem cell; transduction;
KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW gene therapy; gene replacement; genetic defect; thalassemia;
KW Gauchier's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW cytokine; wild-type; ds.
XX
OS Homo sapiens.
XX
FH Key
FH CDS 1..3843
FT Location/Qualifiers
FT /tag= a
FT /product= "Human wild-type MDR-1 protein"
FT replace (531..555, GTT)
FT /tag= b
FT /note= "cDNA sequence of G185V human mutant MDR-1 given
FT in AAZ49332"
XX
```

```
PN WO9961589-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11825.
XX
PR 28-MAY-1998; 98US-0086988.
XX
PA (SUJD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Sorrentino B, Bunting K;
XX
DR WPI: 2000-072615/06.
DR P-FSDB; AAY58186.
XX
PT Ex vivo expansion of hematopoietic stem cells transduced with a
PT sequence encoding human multidrug resistance-1, used for bone marrow
PT transplantation -
XX
PS Claim 10; Page 68-70; 113pp; English.
XX
CC This sequence represents cDNA encoding human wild-type
CC multidrug resistance protein MDR-1. MDR-1 is a transmembrane
CC efflux pump, responsible for the export of drugs from cells,
CC particularly cancer cells. Wild-type MDR-1 shows increased
CC resistance to etoposide and decreased resistance to vinca
CC alkaloids compared with a mutant form (AAY58187) where the Gly at
CC position 185 is replaced by Val. The invention relates to transducing
CC haematopoietic stem cells with nucleic acid encoding an MDR protein
CC and culturing the modified cells. The modified haematopoietic stem
CC cells are useful in bone marrow transplantation (to reconstitute
CC haematopoietic systems in patients who have undergone chemotherapy or
CC radiation therapy) and in ex vivo gene therapy of genetic defects in
CC cells derived from haematopoietic stem cells, e.g., thalassemia,
CC Gauchier's disease, sickle cell anaemia or leukaemia. The modified
CC cells can also be used to identify factors involved in regulating
CC proliferation and differentiation in haematopoietic stem cells.
CC Haematopoietic stem cells that express MDR-1 will be protected against
CC chemotherapeutic agents, so can be engrafted while the patient is
CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC provides sufficient cells to permit standard biochemical analysis.
CC Overexpression of MDR-1 allows cytokine-driven expansion of
CC haematopoietic stem cells by at least 10-fold compared with a maximum
CC of 4-fold in known procedures.
XX
SQ Sequence 3860 BP; 1135 A; 746 C; 957 G; 1022 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 3860
Score: 5859.50 Matches: 1159
Percent Similarity: 95.08% Conservative: 59
Best Local Similarity: 90.48% Mismatches: 60
Query Match: 90.61% Indels: 3
DB: 21 Gaps: 3

US-09-672-725C-4 (1-1280) x AAZ49332 (1-3860)
QY 1 MetAspProGluGlyClyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
|||||
Db 1 ATGGATCTTTGAAGGGGACCGCAATGGAGGAGCAAAAGAAAGAACTTTTAAACTGAAC 60

QY 20 LysLysSerLysLysGluLysLysProThrValSerProThrValSerThrPheAlaMetPhe 39
|||||
Db 61 AATAAAGTGAAGAAAGATAAGAAAGAAAGAAACCACTGTCAGTGTATTTCATGTTT 120

QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle 59
|||||
Db 121 CGCTATTCAAATTCGCTTGACAAGTTGTATATGGTGGGAACTTTGGCGCCATCATC 180

QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
|||||
Db 181 CATGGGGCTGGACTTCCTCTCATGATGCTGCTGTTTGGAGAAATGACAGATATCTTTGCA 240
```


Db 2395 CAGGATGTGAGTTGGTTTGTATGACCCCTAAAAACACCACTGGAGCATTGACTACCAAGGCTC 2454
Qy 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
Db 2455 GCCAATGATGCTCAAGTTTAAGGGGCTATAGGTTCCAGGCTTGCTGTAATTTACCCAG 2514
Qy 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleIleGlyTrpGlnLeuThr 859
Db 2515 AATATAGCAAACTCTGGACAGGAATAATTATATCTTCATCTATGCTGGCACTAACA 2574
Qy 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
Db 2575 CTGTTACTCTTAGCAATGTGTACCATCATTTGCAATAGCAGAGTTGTTGAAATGAAATG 2634
Qy 880 LeuSerGlyGlnAlaLeuLysAspLysGlnLeuGluAlaGlyLysIleAlaThr 899
Db 2635 TTGTCGACAGACACTGAAGATAGAAGAACTAGAAGTGTGGGAAGATCGCTACT 2694
Qy 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
Db 2695 GAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAAAGTTGAACAT 2754
Qy 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
Db 2755 ATGTATCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAGCACACATCTTT 2814
Qy 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
Db 2815 GGAATTACATTTCTTCACCCAGGCAGCATGTATTTTCTTATGCTGGATGTTCCGG 2874
Qy 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 979
Db 2875 TTTGGAGCCTTACTTGGTGGCACATAAACTCATGAGCTTTGAGGATGTTCTGTAGTATT 2934
Qy 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
Db 2935 TCAGCTGTGTTGTTGGTGCATGGCGGTGGGCAAGTAGTCATTTCCTCTGCTGACTAT 2994
Qy 1000 AlaLysAlaLysValSerAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
Db 2995 GCCAAAGCCAAATATACAGAGCCCATCATCATCATCATTTGAAAGAAACCCCTTTGAT 3054
Qy 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
Db 3055 GACAGCTACAGCAGCGAAGCGCTAATGTCGCAACACATCATCATTTGAAAGAAATGCTCATTTGT 3114
Qy 1040 GluValValPheAsnTyrProThrArgProAspIleProValIleGlnIleLeuSerLeu 1059
Db 3115 GAAGTTGATTCNACTATCCACCCGACCGGACATCCAGTGTTCAGGGAGTGTGCTGATGCG 3174
Qy 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
Db 3175 GAGGTGAAGAGGGCCAGACGCTGGCTCTGGTGGCAGAGTGGCTGTGGGAAGAGCAC 3234
Qy 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
Db 3235 GTGGTCCAGCTCTCGAGCGGTCTACGACCCCTTGGCAGGAAAGTGTGCTGATGCG 3294
Qy 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
Db 3295 AAAGAAATAAGCGACTGAATGTTCAAGTGGCTCCGAGCACACCTGGGCATCGTGTCCAG 3354
Qy 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
Db 3355 GAGCCCATCTCTTTGACTGACGATGCTGAGACATGCTGATGGAGACACAGCCGG 3414
Qy 1140 ValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPheIle 1159
Db 3415 GTGGTGTACAGGAAGAGATCGTGGGCGAGCAAGAGGAGGCAACATACATGCTTCATC 3474
Qy 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
Db 3475 GAGTCACCTGCCTAATAATATAGCACTAAGTAGGAGCAAGGAACCTCAGCTCTCTGTT 3534

Qy 1180 GlyGlnLysGlnArgIleAlaIleAlaLeuValArgGlnProHisIleLeuLeu 1199
Db 3535 GGCAGAAACACAGCATTCATAGCTCGTGCCTTGTAGACAGCTCATATTTGCTT 3594
Qy 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
Db 3595 TTGATGAGCCACGTCAGCTCGGATACAGAAAGTGAAGAGTTGTCCAGAAAGCCCTG 3654
Qy 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 3655 GACAAAGCCAGAGAGCCGACCTCATTTGTGATTGCTCACCGCTGTCCACCATCCAG 3714
Qy 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
Db 3715 AATGCAGACTTAATAGTGTGTTTTCAGAAATGCAGAGTCAAGAGCATGGCAGCATCAG 3774
Qy 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 3775 CAGCTCTGGCACAGAAAGCATCTATTTTCAATGGTCAAGTGTCCAGGCTGGAACAAAG 3834
Qy 1280 Arg 1280
Db 3835 CGC 3837

RESULT 7

ABA94365
ID ABA94365 standard; DNA; 3860 BP.
XX
AC ABA94365;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human BCRP DNA related seq Id No. 1.
XX
KW Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
KW hepatotropic; neurodegenerative; cytotstatic; antianemic; muscular; BCRP;
KW cardiant; gene therapy; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..3843
FT /*tag= a
XX
PN WO200192877-A2.
XX
PD 06-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US17459.
XX
PR 31-MAY-2000; 2000US-0584586.
PR 29-MAY-2001; 2001US-0866866.
XX
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Sorrentino B, Schuetz J;
XX
DR WPI; 2002-114368/15.
DR P-PSDB; ABB07266.
XX
PT Identifying a stem cell, for treating e.g., muscular dystrophy,
PT myocardial infarction, Parkinson's disease, or neurodegenerative
PT disorders, comprises detecting the expression of an ATP transport
PT protein (BCRP) by a cell
XX
PS Disclosure; Page 53-55; 87pp; English.

CC The invention provides a method of identifying and/or isolating a stem
CC cell that involves detecting the expression of an ATP transport protein
CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
CC comprising stem cells. The isolated stem cells may be used in the
CC treatment of diseases such as muscular dystrophy, degenerative liver

CC disorder, myocardial infarction, Parkinson's disease, degenerative
 CC disorders of the brain, and for tissue regeneration or replacement.
 CC Haematopoietic cells can be used in bone marrow transplants (e.g., for
 CC treatment of leukemia) and for ex vivo gene therapy for treating blood
 CC diseases such as sickle cell anemia and thalassemia. The stem cells can
 CC also be used as cell targets in gene therapy protocols. The present
 CC sequence represents a sequence related to the BCRP for which no relevant
 CC information has been provided in the specification.
 XX
 SQ Sequence 3860 BP; 1135 A; 746 C; 957 G; 1022 T; 0 other;

Alignment Scores: 0 Length: 3860
 Pred. No.: 5859.50 Matches: 1159
 Score: 95.08% Conservativeness: 59
 Best Local Similarity: 90.48% Mismatches: 60
 Best Local Similarity: 90.61% Indels: 3
 Query Match: 24 Gaps: 3
 DB: 24

US-09-672-725C-4 (1-1280) x ABA94365 (1-3860)

QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPhetRpyLysMetGly 19
 Db 1 ATGGATCTTGAAGGGACCCGCAATGAGGAGCAAGAAAGAACTTTTAAACTGAAC 60
 QY 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
 Db 61 AATAAAGTGAAGATAGAGAGAAAGAAAGAAACCAACTGTCAGTGTATTTCAATGTTT 120
 QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLallelle 59
 Db 121 CGCTATTCAAAATGGCTTGACAAAGTTGTATGTTGGTGGGAACTTTGGCTGCCATCATC 180
 QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspPheAla 79
 Db 181 CATGGGCTGGATCTCTCATGATGCTGCTGTTGGAGAAATGACAGATATCTTTGCA 240
 QY 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleLeuAsnGluSerIleThrAsn 99
 Db 241 AATGCAGGA---AATTTAGAAGATCTGATGTCACAAACATCATTAATAGAAGTATATCAAT 297
 QY 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrThrAlaTyrTyr 119
 Db 298 GATACAGGGTCTTTCATGAAT---CTGGAGAAAGACATACCAAGATGCTTATTTATAC 354
 QY 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
 Db 355 AGTGGAAATGGTGGGCTGCTGCTGCTGCTTACATTCAGGTTCATTTTGGTGGCTG 414
 QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
 Db 415 GCAGCTGGAAGACAAATACACAAATTTAGAAAACAGTTTTTTCATGCTATAATGCGACAG 474
 QY 160 GluIleGlyTrpPheAspValHisAspValGlyLysAsnThrArgLeuThrAspAsp 179
 Db 475 GAGATAGCTGGTGTGATGTCAGATGTTGGGAGCTTAACCCGACATTCAGATGAT 534
 QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
 Db 535 GTCTCTAGATTAATGAAGTATTGGTGGACAAATTTGGAATTTGTTTTCAGTCAATGCA 594
 QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
 Db 595 ACATTTTTCAGTGGGTTTATAGTAGATTTACAGCTGGTGGAAAGCTTAACCCCTTGTGATT 654
 QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
 Db 655 TTGGCCATCAGTCTCTTCTTGGACTGTGACCTGCTGCTGGGCAAGAACTATCTTCA 714
 QY 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
 Db 715 TTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTTTG 774

QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
 Db 775 GCAGCAATTAGAACTGTGATTTGGAGGACAAAGAAAGAACTTTGAAGGTCAAC 834
 QY 280 LysAsnLeuGluAlaLysArgIleGlyLysLysAlaIleThrAlaAsnIleSer 299
 Db 835 AAAAAATTTAGAAAGCTAAAGAAATTTGGATAAAGAAAGCTATTACAGCCCAATTTCT 894
 QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
 Db 895 ATAGTGCTGCTTCTCTGCTGATCATCTATGCTCTGCGCTTCTGCTATGGGACC 954
 QY 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPheSerVal 339
 Db 955 ACCTTGGTCTCTCAGGGGATATTTATTGGCAAGTACTACTGATGATTTCTTCTGTA 1014
 QY 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
 Db 1015 TTAATTTGGGCTTTTAGTGTGGACAGCATCTCCAGCATTTGAAGCATTTGCAATGCA 1074
 QY 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
 Db 1075 AGAGGAGCAGCTTATGAATCTTCAAGATATTTGATAATTAAGCAAGTATTTGACAGTAT 1134
 QY 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
 Db 1135 TCGAAGAGTGGGCAAAACCAAGATATTAAGGAAATTTGGAATTCAGAAATGTTTCA 1194
 QY 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
 Db 1195 TTCAGTTACCCATCTCGAAAAGAAAGTTAAGATCTTGAAGGGCTTGAACCTGAAGTGCAG 1254
 QY 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrValGln 439
 Db 1255 AGTGGCAGACGGTGGCCCTGTTTGGAAACAGTGGCTGGGAAGAGCACAACAGTCCAG 1314
 QY 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
 Db 1315 CTGATGACAGAGCTCTATGACCCACAGAGGGATGGTTCAGTGTGATGACAGATATT 1374
 QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
 Db 1375 AGGACCAATAATGTAAGTCTTACGGGAAATCATTTGGTGGTGGTGGTGGTGGTGGTGGT 1434
 QY 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
 Db 1435 TTGTTGGCCACCACCATAGCTGAAACATTCCTATGCGCCGTGAAATGTCCACATGAT 1494
 QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
 Db 1495 GAGATTGAGAAAGCTCTCAAGGAAGCAATGCTATGACTTTTATCATGAAACGTGCTCAT 1554
 QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
 Db 1555 AAATTTGACACCTGTTGGAGAGAGGGGCCCTGAGTGGTGGTGGGAGAGGAGGAGGAGG 1614
 QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
 Db 1615 ATCGCATTTGACAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1674
 QY 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
 Db 1675 TCAGCTTGGACACAGAAAGCAAGCAGTGTTCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1734
 QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
 Db 1735 GGTGCGACCACTTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1794
 QY 600 AlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
 Db 1795 GCTGCTTTCAGTATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1854
 QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639

KW efflux pump; ss.
XX Macaca fascicularis.
XX
XX
FH Key Location/Qualifiers
FT CDS 100...3942
FT /*tag= a
FT /product= "PGP"
FT /note= "P-glycoprotein"
XX
XX W0200123565-A1.
XX
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26592.
XX
XX 28-SEP-1999; 99US-0156921.
XX 12-OCT-1999; 99US-0158818.
XX
XX (GENT-) GENTEST CORP.
XX
XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
XX WPI; 2001-316136/33.
XX P-PSDB; AAB81064.
XX
XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
XX (PGP) and homologous PGP polypeptides are useful for predicting
XX bioavailability of compound and increasing Pgp transporter activity in
XX cell -
XX
XX Example 1; Page 51-57; 84pp; English.
XX
XX This invention relates to a polynucleotide sequence encoding a
XX cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
XX PGP protein. PGP, also known as multidrug transporter, MDR1 is a member
XX of the ABC transporter superfamily. The enzyme serves as an efflux pump
XX exporting small molecules across the cell membrane. The invention
XX includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
XX and protein, and also that of an allelic variant. The PGP polynucleotide
XX sequence is useful for increasing Pgp transporter activity in a cell.
XX Antisense sequences of the cDNA are useful for inhibiting PGP transport
XX activity in a mammalian cell. They may also be used for increasing the
XX bioavailability of a drug. The present sequence represents cDNA encoding
XX cynomologous monkey P-glycoprotein.
XX
XX SQ Sequence 4186 BP; 1226 A; 799 C; 1039 G; 1122 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 4186
Score: 5855.50 Matches: 1158
Percent Similarity: 95.32% Conservative: 63
Best Local Similarity: 90.40% Mismatches: 57
Query Match: 90.54% Indels: 3
DB: 22 Gaps: 3

US-09-672-725C-4 (1-1280) x AAF86127 (1-4186)

Qy 1 MetAspProGluGlyGlyArgLysGlySerAlaGlu---LysAsnPheTrpLysMetGly 19
Db 100 ATGGATCTTGAAGGGAGCCCAATGGAGGACAGAGAGAACTTTTAAACTGAAC 159
Qy 20 LysLysSerLysLysGluLysGluLysLysProThrValSerThrPheAlaMetPhe 39
Db 160 AATAAAGTAAAGAAAGATAAGAGAAAGAAACCAACTGTGAGTATTTTCAATGTTT 219
Qy 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIlelle 59
Db 220 CGCTATTCAATGCTTGACAAAGTTGTATATGGTGGGGAAGCTTTGGCTGCCATCATC 279
Qy 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
Db 280 CATGGAGCTGGACATTCCTCTCATGATGCTGGTGTGGAGACATGACGGATACCTTTGCA 339

Qy 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
Db 340 AATGCAGGA---AATTAGGAGATTAGGAGCTCTGTTGACTAATAGCAGTAATATCACT 396
Qy 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 397 GATACAGTGGCCGTCATGAAT--CTGGAGGAAGATATGACAGGATATGCTATTATATAC 453
Qy 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
Db 454 AGTGAATTTGGTCTGGGTGCTGGTTCGTCGCTTACATTCAGGTTTCATTTTGGTGCTG 513
Qy 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
Db 514 GCAGCTGGAACACAAATACACAAATTAGAAAACAGTTTTCATGCTATAATGCGACAG 573
Qy 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
Db 574 GAGATAGGCTGGTTTGATGTGCAGATGTTGGGAGCTTAACACCCGCTTACAGATGAT 633
Qy 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
Db 634 GTCCTCAAGATTAAATGAAGGAATTTGGTGACAAAATTTGGAATGTTCTTCAGTCAATGCCA 693
Qy 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
Db 694 ACATTTTTCACCTGGGTTTATAGTAGGATTACACGTGGTTGGAAAGCTAACCCCTTGTGATT 753
Qy 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
Db 754 TTGGCCATCAGTCCTGCTTCTTGGACTGTGAGTGCAGTCTGGCAAGATACTGCTTCA 813
Qy 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu 259
Db 814 TTTACTGATAAGAACTCTTAGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTCTTG 873
Qy 260 AlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyrAsn 279
Db 874 GCAGCAATTAGAACTGTGATTGCAATTTGGAGCACAAGAAAGAACTCGAAAGGTACAC 933
Qy 280 LysAsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
Db 934 AAAATTTAGAGAGAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCAATATTTCT 993
Qy 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
Db 994 ATAGGTGCTGCTTCTGCTTATCTATGCTATGCTTCTGCTTCTGCTATGGGACC 1053
Qy 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
Db 1054 ACCTTGGTCTCTCAAGGAATATCTATTGGACAAGTACTCACTGTATTCTTTTCTGTA 1113
Qy 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
Db 1114 TTAATTGGGGCTTTTAGTGTGGACAGGCATCTCCAAGCATTTGAAGCATTTCCAATGCA 1173
Qy 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
Db 1174 AGAGGAGAGCTTTTGAATCTTCAAGATTAATGATAATAGCCCAAGTATTACAGCTAT 1233
Qy 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
Db 1234 TCAAGAGTGGGCACAAACAGATAATAATAAGGGAATTTTGAATTTCAAGAAATGTTAC 1293
Qy 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
Db 1294 TTCAGTTACCATCTCGAAAGAAAGTAAGATCTTGAAGGGCTGAACTGAACTGAGGTGCA 1353
Qy 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
Db 1354 AGTGGGACAGCGTGCCCTCGCTGGTGGAAACAGCGGCTGTGGGAAGACACACAGGTCCAG 1413

QY 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleaspGlyGlnAspIle 459
Db 1414 CTGATGCAGAGCCTTATGACCCACAGAGGCGCATGTGTGATGACAGGATATT 1473
QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
Db 1474 AGGACCATTAACGTAAGGTTTCTACGGGAATCATCGTGTGTGAGTCAGGAACCTGYA 1533
QY 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
Db 1534 TTGTTTGCACACAGATAGCTGGAACATTCGTATGTGTCGTAAGATGTCACCATGGAT 1593
QY 500 GluIleGluTyrAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
Db 1594 GAGATTGAGAAAGCTGTCAGGAAGCAATGCCTATGACTTTATCATGAACATGCCCTCAG 1653
QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlnLysGlnArg 539
Db 1654 AAATTTGACACCCCTGGTTGGAGAGAGAGGGGCCCGCAGCTGAGTGGTGGCAGAAGCAGAGG 1713
QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
Db 1714 ATCCGCATTGCAGCTGCCCTGGTTGCGCAACCCCAAGATCTCTCTGCTGGACGAGGCCAGG 1773
QY 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
Db 1774 TCAGCCTTGGACACAGAAAGTGAAGCAGTGGTTTCAGTGGCTCGCTGGATAAGGCCAGAAA 1833
QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
Db 1834 GGTGGGACCCACCAATTGTGATAGTCATCGCTTGTCTACCGGTTGCTGAATGCCAGCTCATC 1893
QY 600 AlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
Db 1894 GCTGTTTCATGATGGAGTCATTGTGGAGAAAGGAATCATGATGAGCTCATGAAGAAG 1953
QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
Db 1954 AAAGGCATTTACTTTCAAACTGTGCATATGCAGACAGCAGGAATGAAATGAAATTAGAA 2013
QY 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
Db 2014 AATGCAGCTCATGAATCCAAAGTGAATGTATACCTTGGAAATGCTTCACATGATFCA 2073
QY 660 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
Db 2074 GGATCCAGCTATATAAGAAAAAGATCCACTCGTAGGAGTGTCGGGATCACAGGCCAA 2133
QY 680 AspArgLysLeuGlyLysGluAspLeuAsnGluAsnValProSerValSerPheTrp 699
Db 2134 GACGAAGAGCTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTTTGG 2193
QY 700 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719
Db 2194 AGGATTATCAAGCTAAATTAACAGTGGCCTTATTTTGTGTGTGATTTTGTGCC 2253
QY 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
Db 2254 ATTATAAATGGAGGTGTCACACAGCAATTCGACGTAATATTTTCAAGATATATAGGATT 2313
QY 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
Db 2314 TTTACAGAAATGATGATGCCGCAACAAACGACAGAAATAGTAACCTGTTTCACTATTG 2373
QY 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
Db 2374 TTTCTAGTCTTGGAATTGTTCTTTTATTAATTTTCTTTCAGGCGCTTCACATTGGC 2433
QY 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
Db 2434 AAAGCTGGAGAGATCCTCACCAGCGGCTCGGATACATGTTTTCGATCCATCCTCAGA 2493
QY 800 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819

Db 2494 CAGGATGTGAGCTGGTTTGATGACCTAAAAACACCACTGGAGCATTTGACTACAGGCTC 2553
QY 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
Db 2554 GCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCTATAATTAATACCCAG 2613
QY 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 859
Db 2614 AATATAGCAAACTCTGGACAGGAATAATATATCTTAACTATCTATGTTGGCAACTGACA 2673
QY 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
Db 2674 CTGTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGGAGTTCTTGAATGAAATG 2733
QY 880 LeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGluAlaGlyAlaGlyLysIleAlaThr 899
Db 2734 TTGCTCGACAAGCACTGAAAGATAGAAGAACTAGAGGTTGCTGGGAAGATCGCTACT 2793
QY 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
Db 2794 GAAGCAATAGAAAACCTTCCGAACCTGTTGTTCTTCTGACTCAGGAGCAGAAGTTTGAACAT 2853
QY 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
Db 2854 ATGTATGATCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAACACACATCTTT 2913
QY 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
Db 2914 GGAATCAGCTTTCCITTCAGCAGGCAATGATGATTTTCTCTATGCTGGATGTTCCGG 2973
QY 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 979
Db 2974 TTTGGAGCCTTACTTGTGGCACATAGTCTCATGAGCTTTGAGGATGTTCTGTAGTATT 3033
QY 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
Db 3034 TCAGCTGTTGCTCTTGGTCCATGGCCGTGGGCAAGTCAGTTCAITTTGCTCCTGACTAT 3093
QY 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
Db 3094 GCCAAAAGCCAAAGTATCAGCAGCCACATCATGATCATTTGAAAAAACCCCTTTGATT 3153
QY 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
Db 3154 GACAGCTACGACACAGAGGCTTAAAGCCGAACATTTGGAAGGAATGTCCACATTTAAT 3213
QY 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
Db 3214 GAAATTTGTTTCAACTATATCCACCCGACTGGACATCCCAAGTCTTCAGGGGCTGAGCCTG 3273
QY 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
Db 3274 GAAGTGAAGAGGGCCAGACGCTGGCCCTGGTGGGACAGTCGCTGTGGGAAGACGACG 3333
QY 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
Db 3334 GTGGTCCAGCTCCTCGAGCGGTTCTATACCCCTTGGCGGGGAAGTGTGCTGTGACGGC 3393
QY 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
Db 3394 AAAGAAATAAGCAACTGAATGTTTCAGTGGCTCCGAGCAGACCTGGGCGATCGTCCAG 3453
QY 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
Db 3454 GAGCCCATCTCTGTTTGAAGTGGCATTTAGTAGAACAATGCTGCTATGAGACAACAGCGCG 3513
QY 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1159
Db 3514 GTGGTGTACAGAGACATGCTGAGGCGACCCAGGAGGCAATATACAGCGCTTCATC 3573
QY 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179

Db 742 TGGAAAGCTAACCCCTGTGTGATTTTGGCCATCAGTCCTCTGTCTTGACCTGTCCAGCTGCAGTC 801
QY 233 TrpAlaLysIleLeuSerSerPheThrAspLysGluLeuLeuAlaLysAlaGly 252
Db 802 TGGCAAGATACCTCTTCATTTACGTAAAGAACTCTTAGCTTATGCAAAAGCTGGA 861
QY 253 AlaValAlaGluGluValLeuAlaAlaIleArgThrValIleAlaPheGlyGlyGlnLys 272
Db 862 GCAGTAGCTGAAGAGGCTTGGCAGCAATTAGAACTGTGATTGCAITGGAGGACAAAG 921
QY 273 LysGluLeuGluArgTyrAsnLysAsnLeuGluGluAlaLysArgIleGlyIleLysLys 292
Db 922 AAAGAACTGGAAGGTACACAAATAATTTAGAAGAGCTAAAGAAATTTGGATAAAGAAA 981
QY 293 AlaIleThrAlaAsnIleSerIleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAla 312
Db 982 GCTATTACAGCAATATTTCTATAGTGTCTTTCTCTCAAGGAATATTTCTATTGCACTTAGCT 1041
QY 313 LeuAlaPheTrpTyrGlyThrSerLeuValLeuSerSerGluTyrThrIleGlyGlnVal 332
Db 1042 CTGGCCCTTCGGTATGGGACACCTTGGTCTCTCAAGGAATATTTCTATTGGACAAGTA 1101
QY 333 LeuThrValPhePheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSer 352
Db 1102 CTCACCTGATCTTCTCTGTATTAATTTGGGCTTTTAGTGTGGACAGGCATCTCCAAGC 1161
QY 353 IleGluAlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsn 372
Db 1162 ATTTGAAGCATTTGCAATGCAAGAGGAGCAGCTTTTGAATCTTCAAGATAATTGAAGGAAAT 1221
QY 373 LysProSerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsn 392
Db 1222 AAGCCAAAGTATTGACAGCTATTGGAAGAGTGGGCACAAACCAAGAGATAATTAAAGGAAAT 1281
QY 393 LeuGluPheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeuLys 412
Db 1282 TTGGAATTCAGAAATCTTCATCTCAGTTACCCATCTCGAAAGAAAGTTAAGATCTTGAAG 1341
QY 413 GlyLeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCys 432
Db 1342 GGCTTGAACCTGAAGGTGCAGAGTGGGCAGCGGTGGCCCTGGTTGGAACAGCGGCTGT 1401
QY 433 GlyLysSerThrThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetVal 452
Db 1402 GGAAGAGGACAAACGGCTCAGCTGATGCAGAGGCTTTATGACCCACAGAGGCGATGGTC 1461
QY 453 CysIleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThrGly 472
Db 1462 AGTGTGTGATGGACAGATATTAGGACCATTAACGTAAAGTTCTACGGGAATCATCGGT 1521
QY 473 ValValSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGly 492
Db 1522 TGGGTGAGTCAGGAACCTGTATTGTGTGCCACCATAGCTGAAACATATTCGCTATGGT 1581
QY 493 ArgGluAsnValThrMetAspGluIleGluLysAlaValLysGluAlaAsnAlaTyrAsp 512
Db 1582 CGTAGAGTGTCAACCATGGATGAGATTGAGAAAGCTGTCAAGGAAGCCCAATGGCCTATGAC 1641
QY 513 PheIleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaArgLeu 532
Db 1642 TTTATCATGAACACTGCCCTCAGAAATTTGACACCCTGGTTGGAGAGAGGGGCCACGCTG 1701
QY 533 SerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIle 552
Db 1702 AGTGGTGGCAGAGCAGAGGATCGCCATTGCAGCTGCCCTGGTTGCGCAACCCCAAGATC 1761
QY 553 LeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnVal 572
Db 1762 CTCCTGCTGGAGAGGCCAGCTAGCCCTTGACACACAAAGTGAAGCAGTGGTTCAGGTG 1821
QY 573 AlaLeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThr 592
Db 1822 GCTCTGGATAAGCCAGAAAAGGTTCGACCCACCATTTCTGATAGCTCATCTGTTGTCTACG 1881

QY 593 ValArgAsnAlaAspValIleAlaGlyPheAspGlyValIleValGluLysGlyAsn 612
Db 1882 GTTCGTAATCCGACGTCATCGCTGGTTTCGATGATGGAGTCATTGTGGAGAAAGGAAT 1941
QY 613 HisAspGluLeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArg 632
Db 1942 CATGATGAGCTCATGAAGAGAAAGGCAATTTACTTCAAACTTGTCAATTCGACAGCA 2001
QY 633 GlyAsnGluIleGluLeuGluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeu 652
Db 2002 GGAATGAAATTTGAATTAGAAATGCAGCTGATGAATCCAAAGTCAAATTTGATACCTTG 2061
QY 653 GluMetSerProLysAspSerGlySerSerLeuIleLysArgArgSerThrArgArgSer 672
Db 2062 GAAATGTCTTCACATGATGATTCAGGATCCAGTCTAATAAGAAAAAGATCCACTCGTAGGAT 2121
QY 673 IleHisAlaProGlnGlyGlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsn 692
Db 2122 GTCCGTGCATCACAAAGCCCAAGACAGAAAGCTTAGTACCAAGAGAGGCTCTGGATGAAAGT 2181
QY 693 ValProSerValSerPheTrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPhe 712
Db 2182 ATACCTCCAGTTCTCTTTGGAGGATTATGAAGCTAAATTTAACTGAGTGGCTTATTTT 2241
QY 713 ValValGlyIlePheCysAlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIle 732
Db 2242 GTTGTGGTGATTTTGTGCCATTATAAATGGAGGTCTGCAACACGAGCTTTGCAGTAATA 2301
QY 733 PheSerArgIleIleGlyIlePheThrArgAspGluAspProGluThrLysArgGlnAsn 752
Db 2302 TTTTCAAAGATTATAGGATTTTACAGAAATGATGATGCCAAACAAACACGACAGAA 2361
QY 753 SerAsnMetPheSerValLeuPheLeuValLeuGlyIleIleSerPheIleThrPhePhe 772
Db 2362 AGTAACTGTGTTTTCATATTGTTCTAGTCTTGAATGTTCTTTTATTACATTTTTC 2421
QY 773 LeuGlnGlyPheThrPheGlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMet 792
Db 2422 CTTCAAGGCTTCACATTTGGCAAGCTGGAGAGATCCCTCACCAAGCGCTCCGATACATG 2481
QY 793 ValPheArgSerMetLeuArgGlnAspValSerTrpPheAspProLysAsnThrThr 812
Db 2482 GTTTTCCGATCCATGCTCAGACAGATGTGAGCTGTTGATGACCTTAAACACCACT 2541
QY 813 GlyAlaLeuThrThrArgLeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySer 832
Db 2542 GGAGCAATTGACATCCAGGCTCGCCAATGATGCTCAAGTTAAAGGGGCTATAGGTTC 2601
QY 833 ArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGlyIleIleSerLeu 852
Db 2602 AGGCTGCTTAAATTTACCCAGAAATATAGCAATCTTGGGACAGGAATAATATATCTTA 2661
QY 853 IleTyrGlyTrpGlnLeuThrLeuLeuAlaIleValProIleIleAlaIleAla 872
Db 2662 ATCTATGTTGGCAACTGACACTGTTACTCTTAGCAATTTGATCCCATCATTCGAATAGCA 2721
QY 873 GlyValValGluMetLysMetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGlu 892
Db 2722 GGAGTGTGTAATGAAATGTTGCTCGCAAGCACTGGAAGATAAGAAAGAACTAGAA 2781
QY 893 GlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeuThr 912
Db 2782 GGTGCTGGGAAGATCGCTACTGAGCAATAGAAAACCTTCGAACTGTGTTCTTTGACT 2841
QY 913 ArgGluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSer 932
Db 2842 CAGGAGCAAGATTTGAACATATGTATCATCAGAGTTTGCAGGTTCACCATACAGAACTCT 2901
QY 933 LeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGlnAlaMetMetTyrPhe 952
Db 2902 TTGAGAAAGCACACATCTTTGGAATCACGTTTTTCCCTTCACGAGGCAATGATGATTTT 2961

QY 953 SerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPhe 972
DB 2962 TCCATAGCTGGATGTTCCGTTGGAGCCTACTTGGTGGCACATAGTCTCATGAGCTTT 3021
QY 973 GlnAspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlnVal 992
DB 3022 GAGGATGTTCTGTAGTATTTTCAGCTGTGTGTTGGTGGCCATGGCCGTGGGGAAGTC 3081
QY 993 SerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValIleMetIle 1012
DB 3082 AGTTCATTTGCTCCTGACTATGCGCAAGCCCAAGTATCATGAGCCACATCATCATGATC 3141
QY 1013 IleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLysProAsnThrLeu 1032
DB 3142 ATTGAAAAACCCCTTTGATTGACAGTACAGCACAGAAGGCTAAAGCCGAACACATGTG 3201
QY 1033 GluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIlePro 1052
DB 3202 GAAGGAATGTCACATTTAATGAAGTTGTATTCACTATCCACCCGACTGGNACATCCA 3261
QY 1053 ValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySer 1072
DB 3262 GTGCTTCAGGGCTGAGCCTGGAAGTGAAGAGGCGCAGACGCTGGCCCTGTGTGGCAGC 3321
QY 1073 SerGlyCysGlyLysSerThrValValGlnLeuGluArgPheTyrArgProLeuAla 1092
DB 3322 AGTGGCTGTGGGAAGACACGGTGGTCCAGCTCTGTGAGCGGTCTATGACCCCTTTGGCG 3381
QY 1093 GlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAla 1112
DB 3382 GGGAAAGTGTCTGTGAGCGCAAGAATAAAGCAACTGAATGTCAGTGGCTCCGAGCA 3441
QY 1113 HisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIle 1132
DB 3442 CACCTGGGCATCGTGTCCAGGAGCCATCTGTTGACTGCAGCATTAGTGAGAACATT 3501
QY 1133 AlaTyrGlyAspAsnSerArgValValSerHisGluGluIleMetGlnAlaAlaLysGlu 1152
DB 3502 GCCATGAGAACACAGCGGGTGTGTACAGGAAGAGATCGTGAGGGCAGCCAAAGGAG 3561
QY 1153 AlaAsnIleHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAsp 1172
DB 3562 GCCAATATACAGCCTTCATCGAGTCACTGCCTAATAAATATAGCACAGTAGGAGAC 3621
QY 1173 LysGlyThrGlnLeuSerGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuVal 1192
DB 3622 AAAGGAACCTCAGCTCTCTGTGTGGCCAGAAACAGCATTCGCATAGCTGTCGCCCTGTT 3681
QY 1193 ArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGlu 1212
DB 3682 AGACAGCCTCATATTTTCTTTTGGATGAAGCCACATCAGCTCTGGATACAGAAAGTAA 3741
QY 1213 LysValValGlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIleAla 1232
DB 3742 AAGGTTGTCCAAGAAGCCTCGACAAAGCCAGAGAAGCCGCTACTGCTGATTTGCTGCT 3801
QY 1233 HisArgLeuSerThrIleGlnAsnAlaAspLeuIleValPheGlnAsnGlyLysVal 1252
DB 3802 CACCCGCTGTCCACCATCCAGATCCAGATCAATAGTGGTGTTCAGAAATGGCAGAGTC 3861
QY 1253 LysGluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIle 1272
DB 3862 AAGGAGCAGGCGACATCAGCAGCTGTGGCAGCAAGAAAGCATCTATTTTTCATGTC 3921
QY 1273 SerValGlnAlaGlyValLysArg 1280
DB 3922 AGTGTCAGGCTGGAGCAAGCGC 3945
RESULT 10
AAZ49333
ID AAZ49333 standard; CDNA: 3860 BP.
XX
AC AAZ49333;

XX DT 14-MAR-2000 (first entry)
XX DE Human G185V mutant multidrug resistance-1 (MDR-1) CDNA.
XX KW Multidrug resistance; MDR-1; P-glycoprotein;
KW transmembrane efflux pump; haematopoietic stem cell; transduction;
KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW gene therapy; gene replacement; genetic defect; thalassaemia;
KW Gauchier's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW cytokine; mutant; ds.
XX OS Synthetic.
OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 1..3843
FT /*tag= a
FT /product= "Human G185V mutant MDR-1 protein"
XX PN WO9961589-A2.
XX PD 02-DEC-1999.
XX PF 27-MAY-1999; 99WO-US11825.
XX PR 28-MAY-1998; 98US-0086988.
XX PA (SUUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX PI Sorrentino B, Bunting K;
XX WPI: 2000-072615/06.
XX P-PSDB; AAY58187.
XX Ex vivo expansion of hematopoietic stem cells transduced with a
XX sequence encoding human multidrug resistance-1, used for bone marrow
XX transplantation -
XX Example 1; Page 79-82; 113pp; English.
XX This sequence represents cDNA encoding human G185V mutant multidrug
XX resistance protein MDR-1, where the Gly residue at position 185
XX of the wild-type protein (AAY58186) is replaced by Val. MDR-1 is a
XX transmembrane efflux pump, responsible for the export of drugs from
XX cells, particularly cancer cells. The wild-type MDR-1 shows increased
XX resistance to etoposide and decreased resistance to vinca alkaloids
XX compared with the G185V mutant. The invention relates to transducing
XX haematopoietic stem cells with nucleic acid encoding an MDR protein
XX and culturing the modified cells. The modified haematopoietic stem
XX cells are useful in bone marrow transplantation (to reconstitute
XX haematopoietic systems in patients who have undergone chemotherapy or
XX radiation therapy) and in ex vivo gene therapy of genetic defects in
XX cells derived from haematopoietic stem cells, e.g., thalassaemia,
XX Gauchier's disease, sickle cell anaemia or leukaemia. The modified
XX cells can also be used to identify factors involved in regulating
XX proliferation and differentiation in haematopoietic stem cells.
XX Haematopoietic stem cells that express MDR-1 will be protected against
XX chemotherapeutic agents, so can be engrafted while the patient is
XX undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
XX provides sufficient cells to permit standard biochemical analysis.
XX Overexpression of MDR-1 allows cytokine-driven expansion of
XX haematopoietic stem cells by at least 10-fold compared with a maximum
XX of 4-fold in known procedures.
XX Sequence 3860 BP; 1135 A; 746 C; 956 G; 1023 T; 0 other;
SQ Alignment Scores:
Pred. No.: 0 Length: 3860
Score: 5850.50 Matches: 1158
Percent Similarity: 95.00% Conservatives: 59
Best Local Similarity: 90.40% Mismatches: 61
Query Match: 90.47% Indels: 3

DB: 21 Gaps: 3
US-09-672-725c-4 (1-1280) x AA249333 (1-3860)
QY 1 MetAspProGluGlyGlyArgGlySerAla---GluLysAsnPheTrpLysMetGly 19
DB 1 ATGGATCTTGAAGGGGACCGCAATGGAGGACAAAGAGAAAGAACTTTTAAACTGAAC 60
QY 20 LysLysSerLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
DB 61 AATAAACTGAAAGATACAGAGAAAGAAACCACTGTCAGTGTATTTTCAATGTTT 120
QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 59
DB 121 CGCTATTCAAATTCGCTTGCACAGTTGTATATGCTGGGAACTTGGCTGCCATCATC 180
QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
DB 181 CATGGGCTGGACATCTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTTGCA 240
QY 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
DB 241 AATCGAGGA---AATTAGAGATCTGATGTCACAAATCACTATATAGAAGTGATACAT 297
QY 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
DB 298 GATACAGGGTTCTTCATGAAT---CTGGAGAGACATGACCAGATATGCTTATTATAC 354
QY 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
DB 355 AGTGGAAATGGCTGGGGTGTGGTGTGCTTACATTTCAGGTTTCATTTGGTGCTG 414
QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
DB 415 GCACCTGGAGACAAATACACAAATTAGAAACAGTTTTTTCATGCTATATATGCGAC 474
QY 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
DB 475 GAGATAGCGTGGTTGATGTGCACGATGTTGGGAGCTTAACACCGACTTACAGATGAT 534
QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
DB 535 GTCTCTAAGATTAATGAAGTTATGGTGACAAATGGAAATGTTCTTTCAGTCAATGGCA 594
QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
DB 595 ACATTTTTCAGTGGTTTATAGTAGATTTACACGTGGTGGAGCTAACCCCTTGATG 654
QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
DB 655 TTGGCCATCAGTCTGTCTTGTGACGTGTCAGCTGCTGCTGGGCAAGATACATCTTCA 714
QY 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
DB 715 TTTTACTGATAAAGACTCTTAGCGTATGCAAAAGCTGGAGSCAGTAGCTGAAGAGTCTTG 774
QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
DB 775 GCAGCAATTAGAACTGTGATTCGATTTGGAGGACAAAGAAAGAACTTGAAGGTACAC 834
QY 280 LysAsnLeuGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
DB 835 AAAAATTTAGAAGAAGCTAAAGAAATGGGATAAAGAAAGCTATTACAGCCAAATATTCT 894
QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpGlyThr 319
DB 895 ATAGGTGCTGCTTCTGCTGATCTATGATCTTATGCTGCTGCTTGGTATGGGACC 954
QY 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
DB 955 ACCTTGGTCTCTCAGGGGAATATTCTATTGGACAGTACTCAGTGTATCTTCTTGTA 1014
QY 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359

DB 1015 TTAATTGGGGCTTTTAGTGTGGACAGGCATCTCCAGCATTTGAACATTTGCAATGCA 1074
QY 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
DB 1075 AGAGGAGCAGCTTATGAAATCTTCAAGATAATTTGATAATAAGCAAGTATTGACACGTAT 1134
QY 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
DB 1135 TCGAAGAGTGGGCACAAACAGATATAATAAGGAAATTTGGAATTTGAGAATGTTCCAC 1194
QY 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
DB 1195 TTCAGTTACCATCTCGAAAAGAGTTAGATCTTCAGAGGGCTGAACCTGAAGGTGCAG 1254
QY 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
DB 1255 AGTGGGCAGACGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGACACAACAGTCCAG 1314
QY 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
DB 1315 CTGATGCAGAGGCTCTATGACCCACAGAGGGGATGGTCAGTGTGATGGACAGGATATT 1374
QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
DB 1375 AGGACCATTAATGTAAAGTTTCTACGGGAATCATTTGGTGTGGTGCAGGAACCTGTA 1434
QY 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
DB 1435 TTGTTTGGCCACCAGCATAGCTGAAAACATTCGTATGGCCGTGAAATGTCACCATGGAT 1494
QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
DB 1495 GAGATTGAGAAAGCTCTCAAGGAAGCCAATGCTATGACTTTATCATGAACTGGCTCAT 1554
QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
DB 1555 AAATTTGACACCTGGTTGGAGAGAGAGGGGCCAGTTGAGTGGTGGGAGAGCAGAGG 1614
QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
DB 1615 ATCGCCATTCGACGTGCCCTGGTTGCAACCCCAAGATCCTCCTGCTGGATGAGGCCAG 1674
QY 560 SerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArgLys 579
DB 1675 TCAGCCTTTGGACACAGAAAGCGAGCGTTCAGGTGGCTGCTGGATAAGGCCAGAAAA 1734
QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
DB 1735 GGTGCGACCACTATGATAGCTCATCTGTTGCTACAGTTCTGTAATGCTGACGTCATC 1794
QY 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
DB 1795 GCTGTTTTCGATGATGGAGTCAATTTGGAGAAGGAAATCATGATGAATCATGAAGAG 1854
QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
DB 1855 AAAGCACTTACTTCAAACTTGTCAATGACAGACAGGAAATGAAGTTGGAATTGAA 1914
QY 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
DB 1915 AATGCAGCTGATGAATCAAAAGTGAATTTGATGCTTGGAAATGCTCTTCAAAATGATCA 1974
QY 660 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
DB 1975 AGATCCAGCTCTAATAAGAAAAAGATCAACTCGTAGGAGTGTCCGTGGATGCACAGCC 2034
QY 680 AspArgLysLeuGlyThrLysLysLeuAsnGluAsnValProSerValSerPheTrp 699
DB 2035 GACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCTTTGG 2094
QY 700 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719

Db	2095	AGGATTATGAAGCTAAATTTAACTGAATGGCGCTATTATTTGTTGGTGATATTTTGTGGCC	2155
Qy	720	IleileAsnGlyGlyLeuGlnProAlaPheSerIlellePheSerArgIlelleGlyIle	739
Db	2155	ATTATAAATGGAGGCTGCAACAGCATTTGCAATAATATATTTTCAAAAGATTATAGGGTT	2214
Qy	740	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	759
Db	2215	TTTACAAAGAAATGTGATGATCCTCGAAACAAACAGCAGAAATAGTAATCTGTTTTCACATATTG	2274
Qy	760	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	779
Db	2275	TTTCTAGCCCTTGGAAATATTTCTTTTATACATTTTTCCTTCAAGCTTTCACATTGGC	2334
Qy	780	LysAlaGlyGluIleLeuThrLysArgIleuArgTyrMetValPheArgSerMetLeuArg	799
Db	2335	AAAGCTGGAGAGATCCTCACCAAGCGGCTCCGATACATGTTGTTTCCGATCCATGCTCAGA	2394
Qy	800	GluAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	819
Db	2395	CAGGATGTGAGTTGGTTGATGACCTTAAACACACCATGGAGCATTTGACTACCAGGCTC	2454
Qy	820	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	839
Db	2455	GCCAAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTAATTACCACG	2514
Qy	840	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTyrGlnLeuThr	859
Db	2515	AAATATAGCAAACTCTGGGACAGGAATAATATATCTCTCATCTATGGTTGGCAACTAACA	2574
Qy	860	LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet	879
Db	2575	CTGTTACTCTTACCAATTGTCACCATTTGACCATTTGCAATAGCAGGAGTTGTGAAATGAAAATG	2634
Qy	880	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr	899
Db	2635	TTGTCTGGACAGCACTGAAGATAGAAGAACTAGAAAGGTGCTGGGAAGATCGCTACT	2694
Qy	900	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr	919
Db	2695	GAAGCAATAGAAAACCTCCGAACCGTGTGTTCTTTGACTCAGGAGCAGAAAGTTTGAACAT	2754
Qy	920	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	939
Db	2755	ATGTTATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGAGAAAGCACACATCTTT	2814
Qy	940	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	959
Db	2815	GGAAATACATTTTCCITTCACCCAGGCAATGATGATATTTTCCATGCTGGATGTTCCGG	2874
Qy	960	PheGlyAlaTyrLeuValAlaAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe	979
Db	2875	TTTGGAGCTACTTGTGTGGCACATAAATCATGAGCTTTGAGGATGTCTGTGTAGTATT	2934
Qy	980	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	999
Db	2935	TCAGCTGTGTGCTTGGTCGATGGCGCGTGGGCAAGTCAGTTCATTTTGCCTCGACTAT	2994
Qy	1000	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1019
Db	2995	GCCAAAGCCAAATATACGACGCCACATCATCATGATCATTTGAAAAAACCCCTTTGATT	3054
Qy	1020	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1039
Db	3055	GACAGCTACAGCAGCGGAAGCCCTAAATGCCCAACACATTCGAAGAGAAATGTACATATTGGT	3114
Qy	1040	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1059
Db	3115	GAAGTTGTATTCAACTATVCCACCCAGCCAGCATCCCACTGCTTTCAGGACTGAGCCCTG	3174
Qy	1060	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1079
Db	3175	GAGGTGAAGAAGGCCAGACGCTGGCTCTGGTGGCAGCAGTGGGTGTGGGAAGACACA	3234

Qy	1080	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1099
Db	3235	GTGGTCAGCTCTCTGGAGCGGTCTACGACCCTTGGCAGGAAGTGTCTGTGATGGC	3294
Qy	1100	LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1119
Db	3295	AAAGAAATAAAGCGACTGAATGTCAGTGGCTCCGAGCACACCTGGGCATCGTCCACG	3354
Qy	1120	GluProlLeuLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1139
Db	3355	GAGCCATCTCTTTGACTGCAGCATPTGCTAGAACATTTGCTATGGAGACACAGCCGG	3414
Qy	1140	ValValSerHisGluGluIleMetGlnAlaLaLaLysGluAlaAsnIleHisHisPheIle	1159
Db	3415	GTGGTGTACAGGAAGAGATCGGTGAGGGCAGCAAGAGGAGGCCAACATACATGCCTTCATC	3474
Qy	1160	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1179
Db	3475	GAGTCACTGCCTTAATAATATAGCACTAAAGTAGGAGACAAAGGAACCTCAGCTCTCGGT	3534
Qy	1180	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1199
Db	3535	GCCCAAGAACACAGCATTGCCATAGCTCTGCCCTTGTAGACACCTCATATTTTGCTT	3594
Qy	1200	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1219
Db	3595	TTGGATGAAGCCAGCTCAGCTCTGGATCAGAAAGTGTGAAAGGTTGTCCAAAGAAGCCCTG	3654
Qy	1220	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1239
Db	3655	GACAAAGCCAGAGAAGCCGACCTGCATGTGTGATGCTCACCGCTGTCCACCATCCAG	3714
Qy	1240	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1259
Db	3715	AATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGAGTCAAGGAGCATGGCAGCATCAG	3774
Qy	1260	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys	1279
Db	3775	CAGTGTCTGGCAGACAGAAAGGCATCTATTTTCAATGTCAGTGTCCAGGCTGGACACAAG	3834
Qy	1280	Arg 1280	
Db	3835	CGC 3837	
	RESULT 11		
	ABA94366		
ID	ABA94366 standard; DNA; 3860 BP.		
XX			
AC	ABA94366;		
XX			
DT	26-MAR-2002 (first entry)		
XX			
DE	Human BCRP DNA related seq Id No. 3.		
XX			
KW	Stem cell; ATP transport protein; ATP-binding cassette; antiparkinson		
KW	hepatotropic; neurodegenerative; cytosstatic; antianemic; muscular; BC		
KW	cardiant; gene therapy; ds.		
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..3843	
FT		/*tag= a	
XX			
PN	WO200192877-A2.		
XX			
PD	06-DEC-2001.		
XX			
PF	30-MAY-2001; 2001WO-US17459.		
XX			
PR	31-MAY-2000; 2000US-0584586.		
PR	29-MAY-2001; 2001US-0866866.		

XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 XX Sorrentino B, Schuetz J;
 XX WPI; 2002-114368/15.
 DR P-PSDB; ABB07267.
 XX
 PT Identifying a stem cell, for treating e.g., muscular dystrophy,
 PT myocardial infarction, Parkinson's disease, or neurodegenerative
 PT disorders, comprises detecting the expression of an ATP transport
 PT protein (BCRP) by a cell -
 XX
 PS Disclosure; Page 59-60; 87pp; English.
 XX
 CC The invention provides a method of identifying and/or isolating a stem
 CC cell that involves detecting the expression of an ATP transport protein
 CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
 CC comprising stem cells. The isolated stem cells may be used in the
 CC treatment of diseases such as muscular dystrophy, degenerative liver
 CC disorder, myocardial infarction, Parkinson's disease, degenerative
 CC disorders of the brain, and for tissue regeneration or replacement.
 CC Haematopoietic cells can be used in bone marrow transplants (e.g., for
 CC treatment of leukemia) and for ex vivo gene therapy for treating blood
 CC diseases such as sickle cell anemia and thalassemia. The stem cells can
 CC also be used as cell targets in gene therapy protocols. The present
 CC sequence represents a sequence related to the BCRP for which no relevant
 CC information has been provided in the specification.
 XX
 SQ Sequence 3860 BP; 1135 A; 746 C; 956 G; 1023 T; 0 other;

Alignment Scores:
 Pred. No.: 0 Length: 3860
 Score: 5850.50 Matches: 1158
 Percent Similarity: 95.00% Conservative: 59
 Best Local Similarity: 90.40% Mismatches: 61
 Query Match: 90.47% Indels: 3
 DB: 24 Gaps: 3

US-09-672-725c-4 (1-1280) x ABA94366 (1-3860)

QY 1 MetAspProGluGlyArgLysGlySerAla--GluLysAsnPhetRpyMetGly 19
 DB 1 ATGATCTTGAAGGACCCCAATGAGGAGCAAGAAAGAAAGAACTTTTAAACCTGAAC 60
 QY 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
 DB 61 AATAAAGTGAAGAGTAAAGAGGAAAGAAACCACCTGTCAGTGTATTTCAATGTTT 120
 QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIle 59
 DB 121 CGTATATCAAAATGGCTTGACAAGTTGTATATGGTGGGAACCTTGGCTGCCATCATC 180
 QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
 DB 181 CATGGGGCTGGACTTCTCTCATGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 240
 QY 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
 DB 241 AATGCAGGA---AATTTAGAAGATCTGATGTCACACATCACTAATAGAGTGATATCAAT 297
 QY 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 119
 DB 298 GATACAGGGTTCTTCATGAAT---CTGGAGGAAGACATGACAGATATGCTATATTATAC 354
 QY 120 SerGlyIleGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
 DB 355 AGTGGAAATGGTGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 414
 QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
 DB 415 GCAGCTGGGAAGACAAATACACAAATTTAGAAAACAGTTTTCATGCTATTAATCGGCAC 474

QY 160 GluIleGlyTyrPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
 DB 475 GAGATAGAGCTGGTTTGATGTCACGATGTTGGGAGCTTAACACCGACTTACAGATGAT 534
 QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPheGlnSerIleAla 199
 DB 535 GTCCTAAGATATATGAAGTTATTTGGTGCACAAATTTGGAATGTTCTTCAGTCAATGGCA 594
 QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
 DB 595 ACAATTTTCACTGGTGTATAGTAGGATTTACAGCTGGTTGGAAGCTAACCCCTTGAT 654
 QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
 DB 655 TTGGCCATCAGTCTGCTTGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 714
 QY 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
 DB 715 TTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTACTGAAAGAGTCTTG 774
 QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
 DB 775 GCAGCAATTAGAAGCTGATGATTTGGAGGACAAAGAAAGAACTTCAAGGTACAAAC 834
 QY 280 LysAsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
 DB 835 AAAAATTTAGAAGAGCTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCCAATATTC 894
 QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
 DB 895 ATAGTGCTGCTTCTCTGCTGATCTATGATCTTATGCTCTGGCTCTGGTATGGGACC 954
 QY 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
 DB 955 ACCTTGCTCTCTCAGGGGAATATTTATTTGGACAAAGTACTCAGTGTATCTTTCTGTA 1014
 QY 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
 DB 1015 TTAATTTGGGCTTTTAGTGTGGACAGGATCTCCAAAGCATTTGGAATTTCAAGATTTGCA 1074
 QY 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
 DB 1075 AGAGGAGCAGCTTATGAATCTTCAAGATAATTTGATAATTAAGCCAAAGTATTGACAGCTAT 1134
 QY 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
 DB 1135 TCGAAGAGTGGGCACCAACACAGATAATTTAAGGGGAAATTTGGAATTTCAAGAAATTTCC 1194
 QY 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLysValGln 419
 DB 1195 TTCAGTTACCCATCTCGAAAGAGCTTAAGATCTTGAAGGGCTGGAACCTGAAGGTGGAG 1254
 QY 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
 DB 1255 AGTGGGCAGACGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGAGCACAACAGTCCAG 1314
 QY 440 LeuMetGluArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
 DB 1315 CTGATGCAGAGGCTCTATGACCCACAGAGGGGATGGTCAAGTGTGTGATGGACAGATATT 1374
 QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
 DB 1375 AGGACCAATAATGTAAGGTTTCTACGGGAATCATTTGGTGGTGGTGGTGGTGGTGGTGG 1434
 QY 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
 DB 1435 TTGTTTGGCCACCACTAGCTGAAACATTCGCTATGCCCGTGGAAATGTCCACATGGAT 1494
 QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
 DB 1495 GAGATTGAGAAGAGCTGTCAAGGAAGCCCAATGCCTATGACTTTTATCATGAAACTGCC 1554
 QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539

D	b	1555	AAATTTGACACCCCTGGTGGAGAGAGGGCCCGAGTTGAGTGGTGGCGACAGACGAGG	1614
Q	y	540	IleAlaIleAlaAqAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	559
D	b	1615	ATCGCCATTGACGTGCCCTGGTTCGCAACCCCAAGATCCTCCTGCTGGATGAGGCCAGG	1674
Q	y	560	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaAqLys	579
D	b	1675	TCAGCCTTGACACAGAAACCGACGTGGTTCAGTGGCTCTGGATAGGCCAGCAAA	1734
Q	y	580	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	599
D	b	1735	GGTCGGACCACTTGATGACTCATCTGTGTCTACAGTTCTGTAATGCTGACGTCATC	1794
Q	y	600	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	619
D	b	1795	GCTGGTTTCGATGATGAGTGCATTGTGGAGAAGAAATCATGATGAACTCATGAAGAG	1854
Q	y	620	LysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	639
D	b	1855	AAAGCAATTTACTTCAAACTGTCACAATGCACAGACAGCAAGAAATGAAGTTGAATGAA	1914
Q	y	640	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	659
D	b	1915	AATCAGCTGATGAATCCAAAGAGTGAATGTGATGCTTGGAAATGTCTTCAAATGATTC	1974
Q	y	660	GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln	679
D	b	1975	AGATCCAGTCTAATAGAAAAGAAAGTCAACTCGTAGGAGTGTCCGTGGATCACAGCCCA	2034
Q	y	680	AspArgLysLeuGlyThrLysGluAspLeuAsnValProSerValSerPheTrp	699
D	b	2035	GACAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCCTTTGG	2094
Q	y	700	ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla	719
D	b	2095	AGGATTATGAAGCTTAAATTTAACTGAATGGCCTATTATTGTTGTTGGTATTATTGTGC	2154
Q	y	720	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	739
D	b	2155	ATTATAAATGAGCGCTGCAACAGCAGCATTTGCAATAATATTTTCAAAGATTATAGGGTT	2214
Q	y	740	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	759
D	b	2215	TTTACAGAAATTGATGATCTCGTGAACAAACAGCAGAAATGATGTACTTGTTCATATTG	2274
Q	y	760	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	779
D	b	2275	TTTCTAGCCCTTGGAAATTTTCTTTTATTACATTTTCTCTTCAAGGTTTCACATTTGGC	2334
Q	y	780	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	799
D	b	2335	AAAGCTGGAGAGATCTCCACAGCGCTCCGATATCATGGTTTCCGATPCCATGCTCAGA	2394
Q	y	800	GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	819
D	b	2395	CAGGATGTGAGTTGTTGATGACCCATAAAACACCACCTGGAGCATTGACTACCAAGGCTC	2454
Q	y	820	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	839
D	b	2455	GCCAATGATGCTGCTCAAGTTTAAAGGGGCTATAGTTCACGGCTTGTCTGTAATTACCCAG	2514
Q	y	840	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleThrGlyTrpGlnLeuThr	859
D	b	2515	AATATAGCAAACTTGGGACAGGAATAATTATCTTTCATCTATGTTGGCAACTAACA	2574
Q	y	860	LeuLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLysMet	879
D	b	2575	CTGTACTCTTAGCAATTTGTACCATCATTTGCAATAGCAGGAGTGTCTTGAATGAAATG	2634
Q	y	880	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr	899


```
QY 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
|||||
Db 3775 CAGCTGCTGGCACAAAGGCACTATTTTCAATGGTCACTCCAGGCTGGAACAAG 3834
QY 1280 Arg 1280
|||
Db 3835 CGC 3837
RESULT 12
AAH57442
ID AAH57442 standard; cDNA; 4349 BP.
XX
AC AAH57442;
XX
10-SEP-2001 (first entry)
XX
DE Human intestine cell specific cDNA sequence SEQ ID NO:282.
XX
KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
OS Homo sapiens.
XX
PN W0200132927-A2.
XX
XX 10-MAY-2001.
XX
XX 02-NOV-2000; 2000WO-US30396.
XX
PR 04-NOV-1999; 99US-0163508.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Sornasse T, Seilhamer JJ, Watson GA;
PI
XX WPI; 2001-291057/30.
DR
XX
PT New cell and tissue specific polynucleotides useful for diagnosis,
PT prognosis or monitoring of treatments for disorders where the gene is
PT associated with a cancer, immunopathology or neuropathology -
PS
XX Claim 1; Page 207-208; 327pp; English.
XX
CC AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC sequences (I). (I) can have cytostatic, immunomodulatory and
CC neuroprotective activities, and can be used in gene therapy. (I) and
CC proteins (II) encoded by then are used in high throughput screening
CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC their fragments, immunoglobulins, inhibitors, drug compounds and
CC pharmaceutical agents. Expression of (I) in a sample indicates the
CC differentiation of embryonic stem cells into a tissue selected from
CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC tissues. (I) and (II) are used to produce an expression profile that
CC defines a metabolic or developmental process, treatment, condition,
CC disease or disorder. The gene profile can be used for diagnosis,
CC prognosis or monitoring of treatments and for investigating a
CC predisposition to a disorder where the gene is associated with a
CC cancer, immunopathology or neuropathology.
XX
SQ Sequence 4349 BP; 1304 A; 816 C; 1048 G; 1181 T; 0 other;
```

Alignment Scores:

Pred. No.:	0	Length:	4349
Score:	5850.50	Matches:	1158
Percent Similarity:	95.00%	Conservative:	59
Best Local Similarity:	90.40%	Mismatches:	61
Query Match:	90.47%	Indels:	3
DB:	22	Gaps:	3

|||||
Db 3360 GTGGTCCAGCTCCCTGGAGCGGTTCTACGACCCCTTGCAGGGAAGTGTCTGTGATGGC 3419
QY 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
Db 3420 AAGAAATAAAGCAGCTGAATGTTTCAGTGGCTCCGAGCACACCTGGGCATCGTGTCCAG 3479
QY 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTrpGlyAspAsnSerArg 1139
Db 3480 GAGCCCATCTGTTGACTGCAGCATGCTGAGAACATTCCTATGGAGACACAGCCGG 3539
QY 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1159
Db 3540 GTGGTGTCCACAGGAAGAGATCGTGAGGCGAGCAAGAGGCCCAACATACATCGCTTCATC 3599
QY 1160 ThrLeuProGluLysTrpAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
Db 3600 GAGTCACCTGCTTAATAATATAGCACTAAAGTAGGAGACAAAGGAACCTCAGCTCTGGT 3659
QY 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
Db 3660 GCCCAGAAACACGATTCGCATAGCTCGTGCCCTGTTAGACAGCTCATATTTGCTT 3719
QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
Db 3720 TTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAAGGTTGTCCAGAACCCCTG 3779
QY 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 3780 GACAAACCCAGAGAGCCGACCTGCATGTGATTGTCACCGCCTGTCCACCATCCAG 3839
QY 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
Db 3840 AATGCAGACTTAATAGTGGTGTTCAGATGCGAGATGCGAGATCAAGAGCATGGCACGCATCAG 3899
QY 1260 GlnLeuAlaGlnLysGlyIleTrpPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 3900 CAGCTGCTGCACAGAAAGGCATCTATTTTCAATGGTCAGTGTCCAGCTGGAACAAAAG 3959
QY 1280 Arg 1280
Db 3960 CGC 3962
RESULT 13
AAZ94738
ID AAZ94738 standard; cDNA; 4646 BP.
AC AAZ94738;
XX AAZ94738;
XX 01-AUG-2000 (first entry)
XX Human ATP binding cassette ABCB1 (MDR1) cDNA.
DE ABCB1; ATP binding cassette; human; cholesterol; lipid disorder;
KW atherosclerosis; lipid disorder; dyslipidemia; psoriasis;
KW lupus erythematosus; diagnosis; gene therapy; MDR1;
KW multidrug resistance; chromosome 7q21; ss.
OS Homo sapiens.
XX
XX WO200018912-A2.
XX
XX 06-APR-2000.
XX
XX 21-SEP-1999; 99WO-EP06991.
XX
XX 25-SEP-1998; 98US-0101706.
XX
XX (FARB) BAYER AG.
XX
XX Schmitz G, Klucken J;
XX
XX WPI; 2000-293151/25.

XX Adenosine triphosphate binding proteins useful for identifying agents
PT for treating atherosclerosis and other inflammatory disorders -
XX Claim 9; Page 110-112; 154pp; English.
XX

The present sequence is that of human ATP binding cassette
subfamily B protein ABCB1 cDNA. The cDNA was identified using a
differential display method in which monocytes from peripheral
blood were subjected to macrophage differentiation and cholesterol
loading with acetylated low density lipoproteins and subsequent
deloading with high density lipoprotein (HDL3) to identify
cholesterol sensitive genes. The gene maps to chromosome 7q21
and is also termed MDR1 (multidrug resistance). The invention
provides cholesterol-sensitive ABC genes (see AAZ94734-63). These
genes, and polypeptides encoded by them, can be used for diagnostic
and therapeutic applications, and for biochemical or cell-based
assays to screen for pharmacologically active modulator compounds
useful for the treatment of lipid disorders, atherosclerosis or
other inflammatory diseases such as psoriasis and lupus
erythematosus.

XX SQ Sequence 4646 BP; 1371 A; 892 C; 1129 G; 1254 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 4646
Score: 5850.50 Matches: 1158
Percent Similarity: 95.00% Conservative: 59
Best Local Similarity: 90.40% Mismatches: 61
Query Match: 90.47% Indels: 3
DB: 21 Gaps: 3

US-09-672-725C-4 (1-1280) x AAZ94738 (1-4646)

QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 425 ATGGATCTTGAAGGGGACCCGCAATGGAGGAGCAAGAGAGAACTTTTAAACTGAAC 484
QY 20 LysLysSerLysLysLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
Db 485 AATAAAAGTGAAAGATGAAGGAAAGAAACCAACTGTCTAGTATATTTCAATGTTT 544
QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTrpMetLeuValGlyThrMetAlaIleIle 59
Db 545 CGCTATTCAATGGCTTGACAAAGTTGTATATGGTGGGAACTTTGGCTGCCATCATC 604
QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
Db 605 CATGGGCTGGACTTCCTCTCATGATGCTGTGTTTGGAGAAATGACAGATATCTTTGCA 664
QY 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
Db 665 AATGTCAGGA---AATTTAAGAGATCTGATGTCAAACATCACTAATAAGAAGTATCAAT 721
QY 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTrpAlaTrpTyrTyr 119
Db 722 GATACAGGGTCTTCATGNAAT---CTGAGGAAGACATGACCAGGTATGCCCTATTATTAC 778
QY 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTrpIleGlnValSerPheTrpCysLeu 139
Db 779 AGTGAATTTGGTGTGGGTGCTGTTGCTTACATTCAGGTTTCATTTTGGTGCCTG 838
QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
Db 839 GCAGCTGGAAGACAAATACAAAATTAGAAAACAGTTTTTTCATGCTATATAATGCGACAG 898
QY 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
Db 899 GAGATAGGCTGGTTGATGTGCAGATGTTGGGAGCTTAACACCCGACATTACAGATGAT 958
QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPheGlnSerIleAla 199
Db 959 GTCTCTAAGATTAAATGAAGTTATTGGTGACAAAATTGGAATGTTCTTTCAGTCAATGGCA 1018

Db	3179	ATGTATGCTCAGAGTTTCAGGTACCATACAGAAATCTTTTTCAGGAAAGACACATCTTT	3238
QY	940	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	959
Db	3239	GGAAATTACATTTTCCCTTCACCCAGGCNATGAUGATTTTTTCTATGCTGGATGTTCCGG	3298
QY	960	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe	979
Db	3299	TTTGAGGCTACTTTGGTGGCACAATAACTCAGCTTTGAGCATGTTCTGTAGTATTT	3358
QY	980	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	999
Db	3359	TCAGCTGTGCTTTGGTGCCATGGCCGTGGGGCAAGTCAGTTCATTTGTCCTCGACTAT	3418
QY	1000	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1019
Db	3419	GCCAAAGCCAAATATCAGCAGCCACATCATCATCATGTAATGAAANAACCCCTTTCATT	3478
QY	1020	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1039
Db	3479	GACAGCTACACGACGGAGCCCTAAATGCCGGAACATTTGGAGGAATATGTACATTTGGT	3538
QY	1040	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1059
Db	3539	GAAAGTTGTATTCAACTATCCACCCGACCGACATCCACAGTCCTTCAGGACATGAGCCTG	3598
QY	1060	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1079
Db	3599	GAGGTGAAGAAAGGCCAGACGCTGGCTCTGTGGGCAGCAGTGCCTGTGGAGAGACACA	3658
QY	1080	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1099
Db	3659	GTGCTCCAGCTCCTGGAGCGGTCTACGACCCTTGGCGGAAGTCTGCTGTATGGGC	3718
QY	1100	LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1119
Db	3719	AAAGAAATAAGCGACTGAATGTTCACTGGCTGCCGAGCACACCTTGGCATGCCAG	3778
QY	1120	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1139
Db	3779	GAGCCATCTGTTTCACTGCGACATTCGTGAGAACATTCCTATGAGAACACACGCCGG	3838
QY	1140	ValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPheIle	1159
Db	3839	GTGCTGTTCAGGAAAGAGATCTGAGGCGCAGCAAGGAGGCCACATACATGCCCTTCATC	3898
QY	1160	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1179
Db	3899	GAGTCACTGCCTAATAATATAGCACTTAAAGTAGGAGACAAAGAACTACAGCTCTCTGT	3958
QY	1180	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1199
Db	3959	GGCCAGAAACAACGATTCGCCATAGCTCGTCCCTTCTTACAGACGCTCATATTTTGCTT	4018
QY	1200	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1219
Db	4019	TTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGGTGTCTCAAGAAGCCCTG	4078
QY	1220	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1239
Db	4079	GACAAAGCCAGAGAAGCGCCACCTGCATGTTGATTGCTCACCGCTGTCCACCATCCAG	4138
QY	1240	AsnAlaAspIleuIleValIleValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1259
Db	4139	AATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGATCAAGAGCATGCCACGCATCAG	4198
QY	1260	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys	1279
Db	4199	CAGCTGCTGGCACAAAGGCATCTATTTTCATGTGCTAGTCTCCAGGCTGGAACAAAG	4258
QY	1280	Arg 1280	
		!!!	

Qy	60	HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	79
Db	1997	CATGGGGCTGGAGATCTCCCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTTGCA	2056
Qy	80	AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	99
Db	2057	ATATGTCAGGA--AAATTAGAAGATCTGATGTCAACATCACTAATAGAAGTATATCAAT	2113
Qy	100	AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr	119
Db	2114	GATACAGGGTCTTCTCAAT--CTGGAGGAGACATGATCCAGGATGCTGCTATTATTAC	2170
Qy	120	SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu	139
Db	2171	AGTGGAAATTTGGTCTGGGGTCTGCTGCTCTACATTCAGGTTTCATTTTGGTGCCTG	2230
Qy	140	AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln	159
Db	2231	GCAGCTGGAAGACAAATACACAAAATTAGAAACAGTTTTTTCATGCTATAATATGCCACAG	2290
Qy	160	GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp	179
Db	2291	GAGATAGCTGGTGTGATGTGCACCATGTTGGGAGCTTAACACCCGACTTACAGATGAT	2350
Qy	180	ValSerLysIleAsnGluIleGlyAspLysValGlyMetPheGlnSerIleAla	199
Db	2351	GTCTCTAAGATTAAAGAGTTATTTGTCACAAAATTGGAATGTTCTTTCAGTCAATGGCA	2410
Qy	200	ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle	219
Db	2411	ACATTTTTTCACCTGGGTTTATAGTAGATTTACAGTGGTGGAAAGCTAACCCCTGTGAT	2470
Qy	220	LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleIleTrpAlaLysIleLeuSerSer	239
Db	2471	TTGCCCATCAGTCTCTGCTTGGACTGTGCAGCTGCTCTGGGCCAAAGATACTATCTTCA	2530
Qy	240	PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu	259
Db	2531	TTTACTCATAAAGACTCTTAGCGTATGCMAAGCTGGAGCAGTAGCTGAAGAGGCTTG	2590
Qy	260	AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn	279
Db	2591	GCACCAATTAGAAGTGTGATGGATTTGGAGCACAANAAGAACTCTGAAGGTACAAC	2650
Qy	280	LysAsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer	299
Db	2651	AAAAATTTAGAAGACTAAAGAAATTTGGATAAAGAAAGCTATTACAGCCAATATTCT	2710
Qy	300	IleGlyAlaAlaPheLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr	319
Db	2711	ATAGGTGCTGCTTCTCTGCTGATCATCTTATGCTCTGGCCCTCTGCTATGGGACC	2770
Qy	320	SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal	339
Db	2771	ACCTTGGTCTCTCAGGGGAATTTCTATTGGCAAGTACTCACCTGTATTCTTTCTGTA	2830
Qy	340	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	359
Db	2831	TTAATTTGGGGCTTTAGTGTGGTGGACAGGCATCTCAAGCATTTGAAGCAATTTGCAAAATGCA	2890
Qy	360	ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr	379
Db	2891	AGAGGAGCAGCTTATGAATCTTCAAGATATTGATATAAGCCCAAGTATTGCACAGCTAT	2950
Qy	380	SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	399
Db	2951	TCGAAGAGTGGGCACAAACAGATAATATTAAAGGGAAATTTGGAATTCAGAAATGTTCCAC	3010
Qy	400	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	419
Db	3011	TTCACTTACCATTCTCAAAAGAGTTAAGATCTTGAAGGGCTCTGAAGTGAAGTGCAG	3070

Db	5231	GTGCTGTACAGGAGAGATCGTGGAGGGCAGCAAAAGGAGGCCAACATACATGCCTTCATC	5299
Qy	1160	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1179
Db	5291	GAGTCACTGCCTAATAATATACACTAAAGTAGGAGACAAAGAACTCAGCTCTCTGGT	5350
Qy	1180	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1199
Db	5351	GGCCAGAAAACAAGCGCATTCGCCATGCTCGTGCCTTGTTAGACAGCCTCATATTTTGGTT	5410
Qy	1200	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1219
Db	5411	TTGGATGAACCCAGCTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAGAAGCCCTG	5470
Qy	1220	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1239
Db	5471	GACAAAGCCAGAGAAGCGGCACCTGCATTGTGATTGCTCACCGCTGTCCACCATCCAG	5530
Qy	1240	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1259
Db	5531	ANTGCACACTTAATAGTGTGTTTCAGATGCAGAGTCAAGGAGCATGGCAGCATCAG	5590
Qy	1260	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys	1279
Db	5591	CAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGTGTCAGTGTCCAGGCTGGACAACAAG	5650
Qy	1280	Arg 1280	
Db	5651	CGC 5653	
RESULT	15		
ID	AAZ24041		
ID	AAZ24041	standard; cDNA; 8630 BP.	
AC	AAZ24041;		
XX			
DT	04-FEB-2000	(first entry)	
XX			
DE	Retroviral M4 mdr-1 cDNA.		
KW	Retroviral vector; gag gene; gene therapy; chemotherapeutic agent;		
KW	hematopoietic stem cell transformation; mdr-1; ss.		
XX			
OS	Retrovirus.		
XX			
EP	EP955374-A2.		
XX			
PD	10-NOV-1999.		
XX			
PF	07-MAY-1999; 99EP-0250151.		
XX			
PR	08-MAY-1998; 98DE-1022115.		
XX			
PA	(PETT-) PETTE INST HEINRICH.		
PI	Ostertag W, Baum C, Hildinger M;		
XX			
XX	WPI; 2000-001087/01.		
PT	New retroviral vector containing minimal or no gag gene sequence, for		
PT	use e.g. in gene therapy or cloning		
PS	Disclosure; Page 16-18; 35pp; German.		
XX			
CC	This invention describes a novel retroviral vector (RV) containing a gag		
CC	gene fragment having fewer than 400 bp. RV are used: (i) in gene therapy;		
CC	(ii) for cloning genes; (iii) for (over) expression of proteins or RNAs		
CC	and (iv) for transfection of hematopoietic stem cells (especially to		
CC	impart resistance to chemotherapeutic agents). Because of their reduced		
CC	content of viral genes, RV are very safe (no expression of toxic or		
CC	immunogenic proteins, no recombination with other viruses), have		
CC	increased cloning capacity and express non-viral sequences at a high		
CC	level. This sequence encodes a retroviral M4 mdr-1 protein which is		

CC described in the method of the invention.

XX Sequence 8630 BP; 2235 A; 2011 C; 2187 G; 2197 T; 0 other;

Alignment Scores: Pred. No.: 0 Length: 8630 Score: 5950.50 Matches: 1158 Percent Similarity: 95.00% Conservative: 59 Best Local Similarity: 90.40% Mismatches: 61 Query Match: 90.47% Indels: 3 DB: 21 Gaps: 3

US-09-672-725C-4 (1-1280) x AA224041 (1-8630)

QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
DB 1220 ATGGATCTTGAAGGACCCCAATGAGGAGCAAGAAGAAACCTTTTAAACTGAAC 1279
QY 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
DB 1280 AATAAAGTGAAAAGATAGAAAGGAAGAAACCAACTGTCAGTGTATTTCAATGTTT 1339
QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaLallelle 59
DB 1340 CGCTATTCAAATTTGGCTTGACAAGTTGTATATGTTGGTGGGAACCTTTGGCTGCCATCATC 1399
QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
DB 1400 CATGGGCTGGACTCTCTCATGATGCTGTTTGGAGAAATGACAGATATCTTTGCA 1459
QY 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
DB 1460 AATGCAGCA---AATTTAGAGATCATGCTCAACATCATCTAATAAGAGTGATCAAT 1516
QY 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
DB 1517 GATACAGGGTTCTCATGAAT---CTGGAGGAAGACATGACCAAGATGCTATTATTAC 1573
QY 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
DB 1574 AGTGGAAATTTGGTGGGGTCTGCTGCTTCATTCATTCAGGTTTCATTTTGGTGGCTG 1633
QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
DB 1634 GCAGCTGGAAGACAAATACACAAAATAGAAAACAGTTTTTTCATGCTATAATGCGACAG 1693
QY 160 GluIleGlyTrpPheAspValHisAspValGlyCyluLeuAsnThrArgLeuThrAspAsp 179
DB 1694 GAGATAGGCTGGTTTGATGTGCACGATGTTGGGGAGCTTAACCCGACTTACAGATGAT 1753
QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
DB 1754 GTCCTTAGATTAATGAAGTTATTTGGTGACAAAATTGGATGTTTTCAGTCAATGGCA 1813
QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
DB 1814 ACATTTTTCACGTGGTTTATAGTAGGATTTACACGTGGTGGAGCTTAACCCCTGTGATT 1873
QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
DB 1874 TTGGCCATTCAGTCTGTTTGGACTGTGACGTCTGCTGCTGGGCAAGATCTATCTTCA 1933
QY 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
DB 1934 TTTACTGATAAAGAACTCTTAGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTTCTG 1993
QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlnLysLysGluLeuGluAatqTyrAsn 279
DB 1994 GCAGCAATTAGAACTGTGATTGCAATTTGGAGGACAAAAGAAAGAACTTGAAAGGTACAAC 2053
QY 280 LysAsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
DB 2054 AAAAATTTAGAGAAGCTAAAGAATTTGGGATAAGAAGAGCTATTACAGCCAAATATTCT 2113

QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
DB 2114 ATAGGTGCTGCTTCCCTGCTGATCATGCTTATGCTCTGGCCTTCTGGTATGGACC 2173
QY 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
DB 2174 ACCTTTGGTCTCTCAGGGGAATATTTCTATTTGGACAAGTACTACTGTATTCTTTCTGTA 2233
QY 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
DB 2234 TTAATTTGGGGCTTTTAGTGTGGACAGCATCTCCAGCATTTGAAGCATTTGCAATGCA 2293
QY 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
DB 2294 AGAGGACAGCTTATGAATCTTCAAGATATTTGATAATAAGCCAAAGTATTTGACAGTAT 2353
QY 380 SerLysSerGlyHisLysLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
DB 2354 TCGAAGAGTGGGCACAAACCAAGATAATATTAAAGGAATTTGGAAATTCAGAAATGTTTCA 2413
QY 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
DB 2414 TTCAGTTACCATCTCGAAAAGAGTTAAGATCTTGAAGGCGCTGAACCTGAAGGTGCAG 2473
QY 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
DB 2474 AGTGGCAGACGGTGGCCCTGTTGGAAACAGTGGCTGTGGGAAGAGCACAACTAGTCCAG 2533
QY 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
DB 2534 CTGATGTCAGAGGCTCTATGACCCACAGAGGGATGGTCAGTGTGTGAGCAGAGATATT 2593
QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
DB 2594 AGGACCATAAATGTAAGTTTCTACGGGAATCATTTGCTGTTGGTGGTGCAGTCAGAACCTGTA 2653
QY 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
DB 2654 TTGTTTGGCCACCAAGATAGCTGAAACCATTCGCTATGGCCGTGAAATGTCCCATGGAT 2713
QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
DB 2714 GAGATTGAGAAAGCTGTCAAGGAAGCCAAATGCCTATGACTTTATCATCAAACTGCCTCAT 2773
QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
DB 2774 AAATTTGACACCTGTTGGAGAGAGAGGGGCCAGTTGAGTGGTGGCAGAAAGCAGAGG 2833
QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
DB 2834 ATCGCCATTTGCACGTGCCCTGTTTCGAAACCCCAAGATCCCTCTCTGGATGAGGCCACG 2893
QY 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
DB 2894 TCAGCCTTGGACACAGAAAGCGAAGCAGTGGTTTCAAGTGGCTCTGATAAGGCCAGAAA 2953
QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
DB 2954 GGTGGGACCACTTGTGATAGCTCATCGTTTTGTTACAGTTTCTGTAAGTCTGACCTCATC 3013
QY 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
DB 3014 GCTGGTTTCTGATGATGGAGTCAATTTGGGAGAAAGGAAATCATGATCACTCAATGAAGAG 3073
QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
DB 3074 AAAGGCATTTACTTCAAACTTGTACAATGCAGACAGCAAGAAATGAACTGAATTAGAA 3133
QY 640 AsnAlaThrGlyCyluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
DB 3134 AATGTCAGCTGATGAATCCAAAAGTGAATTTGATGCCCTGGAAATGTCTTCAATGATTCA 3193

QY	660	GlySerSerLeuIleLysLeuAspGluAspProGluThrLysArgSerThrArgSerIleHisAlaProGlnGlyGln	679
DB	3194	AGATCCAGCTCTAATAAGAAAAAGATCAACTCGGTAGGAGTTCGCTGGATCAACAAGCCCA	3253
QY	680	AspArgLysLeuGlyThrLysGluAspLysLeuAsnGluAsnValProSerValSerPheThrP	699
DB	3254	GACACAAAGCCTAGTACCAAGAGGCTCTGGATGAAGAGTATACCTCCAGATTTCCTTTTGG	3313
QY	700	ArgIleLeuLysLeuAsnSerThrGluThrProTyrPheValValClyIlePheCysAla	719
DB	3314	AGGATTATGAAGCTAAATTAACCTGAATGCCCTATTATTTGTTGGTGTATTGTGTC	3373
QY	720	IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	739
DB	3374	ATTATTAATGAGGCCCTGCACACAGCATTTGGCAATAAATATTTTCAAAGATTATAGGGTT	3433
QY	740	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	759
DB	3434	TTTACAAGATTGATGATCTCTGAACAAACACAGACAGATAGTAACCTGTTTCACTATTG	3493
QY	760	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	779
DB	3494	TTTCTAGCCCTTGGAAATATTCTTTATTAACATTTTCTCTCAAGGTTTTCACATTTGGC	3553
QY	780	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	799
DB	3554	AAAGCTGAGAGATCTCTCAACAGCGCTCCGATACATGGTTTCCGATCCATCCTCAGA	3613
QY	800	GlnAspValSerThrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	819
DB	3614	CAGAGTGTGATGGTGGTGTGATGACCTAAAAACACACCTGGAGCATTTGACTACCAAGCTC	3673
QY	820	AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	839
DB	3674	GCCAAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCTGTATTTACCCAG	3733
QY	840	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTyrGlnLeuThr	859
DB	3734	AATATAGCAATCTGGGACAGAAATATATATCTTCATCTATGTTGGCAACTAACCA	3793
QY	860	LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet	879
DB	3794	CTGTTACTCTTACCAATTGTCCCATCATTCGATAGCGAGGTGTGAAATGAAGAAATG	3853
QY	880	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr	899
DB	3854	TTGCTGGACAGCACTGAAGATAGAAGAACTAGAAAGGTGCTGGGAAGATCGCTACT	3913
QY	900	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr	919
DB	3914	GAAGCAATAGAAAATTCGAAACCGTGTGTTCTTGACTCAGGACGAGAAGTTTGAACT	3973
QY	920	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	939
DB	3974	ATGTATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTTGAGGAAGACACATCTTT	4033
QY	940	GlyValSerPheSerIlePheGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	959
DB	4034	GGAAATACATTTTCTTACCACGCAATGAATGATTTTTCCTATGCTGGATGTTTCCGG	4093
QY	960	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe	979
DB	4094	TTTGGAGCCTACTTGGTGGCACATAAATCATGAGCTTGAGGATGTTCTGTGTAGTATTT	4153
QY	980	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	999
DB	4154	TCAGCTGTGTGTTTGGGTGCATGGCCGTGGGGCAAGTCAGTTGCTCTCTGCTGACTAT	4213
QY	1000	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1019
DB	4214	GCCAAAGCCAAATATATCAGACGCCCATCATCATGATCATTAATGAAATAAATCCCTTGAT	4273
QY	1020	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1039

Db	4274	GACAGCTACACGACGGAAGGCCTAATGTCGGAACACATTGGAGGAAATGTCCACATTTGGT	4333
QY	1040	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1059
Db	4334	GAAGTTGTATTCAACTATCCACCACCGACGACATCCCAAGTCTTCAGGACTGAGCCTG	4393
QY	1060	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1079
Db	4394	GAGGTGAAGAAGGCCAGACGCTGGCTCTGTGGCGACAGTGGCTGTGGGAAGAGCACA	4453
QY	1080	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1099
Db	4454	GTGTTCCAGCTCTCTGGAGCGGTTCTACGACCCCTTGGCAGGGAAGTCTGCTTGATGGC	4513
QY	1100	LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1119
Db	4514	AAAGAAATTAAGCGACTGAATGTTCACTGGCTCCGAGCAGCACCTGGGCATCGTCCACG	4573
QY	1120	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1139
Db	4574	GAGCCCATCTGTTTGACTGCAGCATGCTGAGACATTTGCTATGGAGACACGCGCG	4633
QY	1140	ValValSerHisGluIleMetGlnAlaLysGluAlaAsnIleHisHisPheIle	1159
Db	4634	GTGTTGTACAGGAAGAGATCGTGGGGCAGCAAGAGGAGGCCAACATACATGCCTTCATC	4693
QY	1160	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1179
Db	4694	GAGTCACTGCCTAATAATATAGCACTAAGTAGGAGACAAAGAACTCAGCTCTCTGGT	4753
QY	1180	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1199
Db	4754	GGCCAGAAACAACGCATTCGCATAGCTCGTGCCCTTGTGTAGACAGCCTCATATTTGCTT	4813
QY	1200	LeuAspGluAlaThrSerAlaLeuaspThrGluSerGluLysValValGlnGluAlaLeu	1219
Db	4814	TTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCGAAGAGCCCTG	4873
QY	1220	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1239
Db	4874	GACAAAGCAGAGAAGCCGCACCTGCATTTGTATGCTCACCCGCTGTCCACATCCAG	4933
QY	1240	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1259
Db	4934	AATGCAGACTTAATAGTTGTTTTCAGAAATGGCAGAGTCAAGGAGCATGCCACGCATCAG	4993
QY	1260	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys	1279
Db	4994	CAGCTGCTGGCACAGAAAGGCATCTATTTTTCATATGGTTCAGTGTCCAGGCTGGAAACAAG	5053
QY	1280	Arg 1280	
Db	5054	CGC 5056	

Search completed: November 6, 2002, 19:11:14
Job time : 592.41 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:39:50 : Search time 58,5103 Seconds
(without alignments)
5373.600 Million cell updates/sec

Title: US-09-672-725C-4

Perfect score: 6467

Sequence: 1 MDPEGRKGSAGNEKFWKMGK.....LLAQKGIYFSMISVQAGAKR 1280

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09672725/runat_04112002_124339_17887/app_query.fasta_1.6435
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725.cgn_1_134 @runat_04112002_124339_17887 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCFUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	5850.5	90.5	4646	1 US-08-181-471-2
2	5850.5	90.5	6505	2 US-08-793-610-5
3	5850.5	90.5	9318	2 US-08-793-610-6
4	5847.5	90.4	4669	6 5206352-3
5	5823.5	90.0	4669	2 US-08-752-447-1
6	5809	89.8	4264	2 US-08-784-649A-1
7	5809	89.8	4264	2 US-08-784-649A-5
8	5790.5	89.5	4669	3 US-08-583-276-18
9	5296	81.9	4233	3 US-09-120-513-1
10	5296	81.9	4233	1 US-09-450-105-1
11	3615	55.9	2726	1 US-08-461-823-1
12	2511.5	38.8	4047	2 US-08-612-734B-1

13	2482	38.4	4002	2	US-08-996-545-1	Sequence 1, Appli
14	2482	38.4	4002	2	US-08-996-545-3	Sequence 3, Appli
15	2482	38.4	4002	4	US-09-328-320-1	Sequence 1, Appli
16	2482	38.4	4002	4	US-09-328-320-3	Sequence 3, Appli
17	2437	37.7	4800	2	US-08-612-734B-3	Sequence 3, Appli
18	2432	37.6	4224	1	US-08-612-521-1	Sequence 1, Appli
19	2267	35.1	3924	1	US-08-995-246C-1	Sequence 1, Appli
20	2046.5	31.6	6143	1	US-08-612-521-3	Sequence 3, Appli
21	1881	29.1	3924	2	US-08-996-644-3	Sequence 3, Appli
22	1881	29.1	3924	3	US-09-352-552-3	Sequence 3, Appli
23	1881	29.1	3927	2	US-08-996-644-1	Sequence 1, Appli
24	1881	29.1	3927	3	US-09-352-552-1	Sequence 1, Appli
25	1860.5	28.8	3909	1	US-08-232-537-1	Sequence 1, Appli
26	1082.5	16.7	4403765	4	US-09-103-840A-2	Sequence 2, Appli
27	1082.5	16.7	4411529	4	US-09-103-840A-1	Sequence 1, Appli
28	783.5	12.1	2376	1	US-08-394-880B-1	Sequence 1, Appli
29	778	12.0	4781	2	US-09-001-273-1	Sequence 1, Appli
30	778	12.0	4781	4	US-08-843-459A-1	Sequence 1, Appli
31	778	12.0	4847	3	US-09-061-400-1	Sequence 5, Appli
32	757.5	11.7	5889	1	US-08-463-092B-5	Sequence 5, Appli
33	757.5	11.7	5889	2	US-08-462-109A-5	Sequence 5, Appli
34	757.5	11.7	5889	2	US-08-460-907B-5	Sequence 5, Appli
35	757.5	11.7	5889	3	US-08-463-179A-5	Sequence 5, Appli
36	757.5	11.7	5889	3	US-08-461-384B-5	Sequence 5, Appli
37	726.5	11.2	6140	4	US-09-439-313-536	Sequence 536, App
38	711	11.0	5232	4	US-08-972-927-1	Sequence 1, Appli
39	708.5	11.0	6082	4	US-09-439-313-535	Sequence 535, App
40	692	10.7	2061	4	US-09-061-764A-17	Sequence 17, Appli
41	690	10.7	5011	1	US-08-463-092B-3	Sequence 3, Appli
42	690	10.7	5011	2	US-08-462-109A-3	Sequence 3, Appli
43	690	10.7	5011	2	US-08-460-907B-3	Sequence 3, Appli
44	690	10.7	5011	3	US-08-463-179A-3	Sequence 3, Appli
45	690	10.7	5011	3	US-08-461-384B-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-181-471-2
; Sequence 2, Application US/08181471
; Patent No. 5641508
; GENERAL INFORMATION:
; APPLICANT: Li, Lingna
; APPLICANT: Lishko, Valeryi K.
; TITLE OF INVENTION: METHOD FOR DELIVERING BENEFICIAL
; TITLE OF INVENTION: COMPOSITIONS TO HAIR FOLLICLES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas Fitting
; STREET: 12526 High Bluff Drive, Suite 300
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92130
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/181.471
; FILING DATE: 13-JAN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/041.553
; FILING DATE: 02-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: ANT0029P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-792-3680
; TELEFAX: 619-792-8477

; INFORMATION FOR SEQ ID NO: 2:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 4646 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: cDNA			
; HYPOTHETICAL: NO			
; ANTI-SENSE: NO			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 425..4267			
; US-08-181-471-2			
Alignment Scores:			
Pred. No.:	0	Length:	4646
Score:	5850.50	Matches:	1158
Percent Similarity:	95.00%	Conservative:	59
Best Local Similarity:	90.40%	Mismatches:	61
Query Match:	90.47%	Indels:	3
DB:	1	Gaps:	3
US-09-672-725c-4 (1-1280) x US-08-181-471-2 (1-4646)			
QY	1	MetAspProGluGlyArgLysGlySerAla---	GluLysAsnPheTrpLysMetGly 19
DB	425	ATGGATCTTGAAGGGGACCCCAATGGAGGAGCAAGAGAAAGAACTTTTAAACTGAAC	484
QY	20	LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe	39
DB	485	AATAAAGTGAAGAAGATAAGAGGAAAGAAACCAACTGTCAGTGTATTTCAATGTTT	544
QY	40	ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle	59
DB	545	CGCTATTCAAAATTGGCTTGCAAGTTGTATATGTTGGTGGGAACTTTGGTGCCATC	604
QY	60	HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	79
DB	605	CATGGGCTGGACTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTTGCA	664
QY	80	AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	99
DB	665	AATGCAGGA---AATTAGAAGATCTGATGTCAAACATCACTAATAAGAGTATATCAAT	721
QY	100	AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlaTyrTyr	119
DB	722	GATACAGGTCTTCATGAAAT---CTGGAGAAGACATGACCAGGTATGCCATTATATAC	778
QY	120	SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu	139
DB	779	AGTGAATTGGTGTGGGGTCTGGTGTGCTTACATTACAGGTTCATTTTGGTGGCTG	838
QY	140	AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln	159
DB	839	CGACGTGAAGACAAATACAAATATAGAAACAGTTTTTTCATGCTAATAATCGGCACAG	898
QY	160	GluIleGlyTyrPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp	179
DB	899	GAGATAGCTGTTTGTATGTCACAGATGTTGGGAGCTTAACACCCGACTTACAGATGAT	958
QY	180	ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla	199
DB	959	GTCTCTAAGATTATAGAAAGTTATTGGTGACAAATATGGAATGTTCTTTCAGTCAATGGCA	1018
QY	200	ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle	219
DB	1019	ACATTTTTCATCGGTTTATAGTAGGATTTACACGTGGTGGAGCTTAACCCCTTGAT	1078
QY	220	LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer	239
DB	1079	TTGGCCATCAGTCTCTTGGAGTGTGAGTGTGCTGGGCAAGATATCTATCTCA	1138
QY	240	PheThrAspLysLysLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu	259

Db 2219 GCTGGTTTCGATGAGTCAATGTGGAGAAAGAAATCATGATCAACTCATGAAGAG 2278
Qy 620 LysGlyIleYrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
Db 2279 AAAGCATTTACTTCAAACTTGTCAATGCACAGACAGGAGAAATGAAGTTGAATTAGAA 2338
Qy 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
Db 2339 AATGCAGTCATGAATCCAAAGTGAATTTGATGCTTGGAAATGCTTCAATGATCA 2398
Qy 660 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
Db 2399 AGATCCAGTCTAATAAGAAAGATCACTCGTAGAGTGTCCGTGATCACAAGCCAA 2458
Qy 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheIrp 699
Db 2459 GACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGTATACCTCCAGTTTCTTTGG 2518
Qy 700 ArgIleLeuLysLeuAsnSerThrGluIrpProTyrPheValIleGlyIlePheCysAla 719
Db 2519 AGGATTATGAAGCTAAATTTAACTGAATGGCTTATTTTGTGTGTGCTATTTTGCC 2578
Qy 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
Db 2579 ATTATAAATGGAGGCTGCAACAGCATTTGCAATAATATTTTCAAGATTATAGGGTT 2638
Qy 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
Db 2639 TTTTACAAGATTGATGATCTCTGAAACAAACACAGATAGTAATCTTTTCTACTATTG 2698
Qy 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
Db 2699 TTTTCTAGCCTTGGATTAATTTCTTTTATTACATTTTTCCTTCAGGGTTTCACATTGGC 2758
Qy 780 LysAlaGlyClnIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
Db 2759 AAACCTGGAGAGATCTCACCAAGGGCTCCGATACATGGTTTCCGATCCATCCTCAGA 2818
Qy 800 GlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
Db 2819 CAGGATGTGAGTTGGTTGATGACCTTAAACACACCATGGAGCATTTGACACCAGCTC 2878
Qy 820 AlaAsnAspAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
Db 2879 GCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTGCTGTAATTAACCCAG 2938
Qy 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTyrGlnLeuThr 859
Db 2939 AATATAGCAATCTTGGACAGGAATAATATATCTTCTATGTTGGCAACTAACA 2998
Qy 860 LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
Db 2999 CTGTTACTCTTAGCAATTTGACCATCTTCAATAGCAGGAGTGTGTAATGAATAATG 3058
Qy 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGlyAlaGlyLysIleAlaThr 899
Db 3059 TTGCTGGACAGCACTCAAGAGATAAGAAAGTCTAGAAAGTGTGGGAAAGATCGCTACT 3118
Qy 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
Db 3119 GAAGCAATAGAAACTCCGACCGTGTGTTTCTTTGACTCAGGAGAGAAAGTTGAACAT 3178
Qy 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
Db 3179 ATGTATGCTCAGAGTTTTCAGGTACCATACAGAAACTCTTTGAGGAAAGCACATCTTT 3238
Qy 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
Db 3239 GGAATACATTTTCTTCAACCGCAATGATGATTTTCTTATGCTGATGATGTTTCGG 3298
Qy 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 979
Db 3299 TTTGGAGCCTACTTGTGGCACATAACTCATGAGCTTTGAGGATGTTCTGTGTAGTATT 3358

Qy 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
Db 3359 TCAGCTTGTCTTGTGTGTCATGGCGTGGGCAAGTCAGTTCATTCTCTCTGACATAT 3418
Qy 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
Db 3419 GCCAAAGCCAAATATACAGAGCCCATCATCATCATCATCATGAAACCCCTTTGATT 3478
Qy 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
Db 3479 GACAGCTACAGCAGCGAGGCTTAATGCCGACACACATTTGGAAGAAATGTCACATTTGGT 3538
Qy 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
Db 3539 GAAGTTGTATTCAACTATCCACCCGACGGACATCCAGTCTTCAGGACTGAGCCTG 3598
Qy 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
Db 3599 GAGGTGAAGAGGCCAGACGCTGGCTCTGGTGGGACAGTGGCTGTGGGAAGAGCACA 3658
Qy 1080 ValValClnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
Db 3659 GTGCTCCAGCTCCTGGAGCGGTTCTACGACCCCTTGGCAGGGAAAGTCTGCTTGTATGGC 3718
Qy 1100 LysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSerGln 1119
Db 3719 AAAGAAATAAAGCGACTGAATGTTCACTGGCTCCGAGCACACCTGGGCATCGTGTCCAG 3778
Qy 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
Db 3779 GAGCCCATCTCTGTTGACTGACAGCATGCTGAGAACATTTGCTATGGAGACAACAGCCGG 3838
Qy 1140 ValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPheIle 1159
Db 3839 GTGGTGTACAGAGAGAGATCGTGAGGCGACAAAGAGGCGCAACATACATGCTTCATC 3898
Qy 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
Db 3899 GAGTCACTCCCTTAATAATATATAGCACTAAAGTAGAGACAAGAACTCAGCTCTCTGGT 3958
Qy 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
Db 3959 GGCCAGAAACAACGATTTGCCATAGCTCGTCCCTTTGTAGACAGCTCATATTTTCTT 4018
Qy 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
Db 4019 TTGGATGAAGCCACCTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAGAGCCCTG 4078
Qy 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 4079 GACAAGCCAGAAAGGCGGACCTGTCATTTGTGATGCTCACCCCTGTCCACCATCCAG 4138
Qy 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
Db 4139 AATGCAGACTTAATAGTGGTGTTCAGAAATGGCAGATCAAGAGCATGGCAGCATCAG 4198
Qy 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 4199 CAGCTGCTGGCACAGAAAGGATCTATTTTCAATGCTCAGTGTCCAGGCTGGAACAAAG 4258
Qy 1280 Arg 1280
Db 4259 CGC 4261
RESULT 2
US-08-793-610-5
; Sequence 5, Application US/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARBERS, Carol
; APPLICANT: OSTERTAG, Wolfram

```

; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
; TITLE OF INVENTION: FOR GENE TRANSFER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,610
; FILING DATE: 07-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 31 973.8
; FILING DATE: 08-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 195 03 952.1
; FILING DATE: 07-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/03175
; FILING DATE: 10-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berman, Richard J.
; REGISTRATION NUMBER: 39,105
; REFERENCE/DOCKET NUMBER: P1614-7007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6505 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
; US-08-793-610-5

Alignment Scores:
Pred. No.: 0 Length: 6505
Score: 5850.50 Matches: 1158
Percent Similarity: 95.00% Conservative: 59
Best Local Similarity: 90.40% Mismatches: 61
Query Match: 90.47% Indels: 3
DB: 2 Gaps: 3

US-09-672-725c-4 (1-1280) x US-08-793-610-5 (1-6505)
Qy 1 MetaspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1817 ATGGATCTTGAAGGGGCCCAATGGAGGAGCAAGAAAGAAAGAACTTTTAAACCTGAAC 1876
Qy 20 LysLysSerLysGluLysLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 39
Db 1877 AATAAAGTGAAGAGATAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1936
Qy 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaAlaAla 59
Db 1937 CGCTATTCAAAATGGCTTGACAAGTTGATATGTTGGTGGGAACTTTGGCTGCCATCATC 1996
Qy 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
Db 1997 CATGGGGCTGCAGCTCTCTCATGATGCTGTGTGTGGAGAAATGACAGATATCTTTGCA 2056
Qy 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleLeuAsnGluSerIleThrAsn 99
Db 2057 AATGCAGGA---ATTTAGAGAGATCTGATGTCACAAACATCCTACTAATAGAGTATATCAAT 2113

```

```

Qy 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 119
Db 2114 GATACAGGGTCTTCATGAAT---CTGGAGGAAGACATGACAGGTATGCTATTATTATAC 2170
Qy 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
Db 2171 AGTGGAAATGGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2230
Qy 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
Db 2231 GCAGCTGGAAGACAAATACACAAAATTAGAAAACAGTTTTTTCATGCTATAATGCGACAG 2290
Qy 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
Db 2291 GAGATAGGCTGGTTGATGTGCAGATGTTGGGAGCTTAACACCCGACTTACAGATGAT 2350
Qy 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
Db 2351 GTCTTAAGATTAAATGAAGTTATTGGTGACAAAATTGGGAATGTTCTTCAGTCAATGGCA 2410
Qy 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
Db 2411 ACATTTTTCAGCTGGTTTATAGTAGGATTACACCTGGTTGGAAAGCTTAACCCCTGTGATT 2470
Qy 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
Db 2471 TTGGCCATCAGTCTCTTGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2530
Qy 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
Db 2531 TTTACTGATAAAGAACTCTTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGTCTTG 2590
Qy 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluAaGlyTyrAsn 279
Db 2591 GCAGCAATTAGAACTGTGATTGCATTTGGAGGACAAAGAAAGAAAGAAAGTGAAGAGTACAAC 2650
Qy 280 LysAsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
Db 2651 AAAAATTTAGAGAGAGCTTAAAGAAATTTGGGATAAAGAAAGCTATTACAGCCAATATTTCT 2710
Qy 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
Db 2711 ATAGTGTGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2770
Qy 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
Db 2771 ACCTTGGTCTCTCAGGGGAATATTTCTATTGGACAGACTACTGATGATTTCTTTCTGTA 2830
Qy 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
Db 2831 TTAATTTGGGCTTTTGTGTTGGACAGGCATCTCCCAAGCATTTGAAGCATTTTGCATAATGCA 2890
Qy 360 ArgGlyAlaAlaTyrGluIlePheLysIleLeuAspAsnLysProSerIleAspSerTyr 379
Db 2891 AGAGAGCAGCTTATGAATCTTCAAGATAATTTGATAATAAGCCCAAGTATTGACAGCTAT 2950
Qy 380 SerLysSerGlyHisLysProAspAsnLysLysGlyAsnLeuGluPheLysAsnValHis 399
Db 2951 TCGAAGAGTGGGCACAAACCAAGATATATTAAAGGAATTTTGAATTCAGAAATGTTCCAC 3010
Qy 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
Db 3011 TTCAGTTACCATCTCGAAAAGAAAGTAAAGATCTTGAAGGGCTTGAAGCTGAAGGTGCAG 3070
Qy 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
Db 3071 AGTGGCAGACGGTGGCCCTGTTGGAAACAGTGGCTGGGGAAGAGGACACACAGTCCAG 3130
Qy 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
Db 3131 CTGATGACAGAGGCTCTATGACCCACAGAGGGGATGGTTCAGTGTGTTGATGACAGGATATT 3190
Qy 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479

```


QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
|||||
Db 5411 TTGGATGAAGCCAGCTCAGCTTGGATACAGAAAGTGAAGGTTGTCCAGAACCCCTG 5470
QY 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
|||||
Db 5471 GACAAAGCCAGAGAGCCGACCTGCATTGTGATGCTCACCCGCTGTCACCATCCAG 5530
QY 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
|||||
Db 5531 AATGCAGACTTAATAGTGTGTTCAGATGGCAGATGCAAGAGCATGGCAGCATCCAG 5590
QY 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
|||||
Db 5591 CAGCTGCTGGCAGAGAGGCACTATTTTCAATGGTCAGTGTCCAGGCTGGACAGAG 5650
QY 1280 Arg 1280
|||
Db 5651 CGC 5653

RESULT 3
US-08-793-610-6
; Sequence 6, Application US/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARRERS, Carol
; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
; NUMBER OF INVENTION: FOR GENE TRANSFER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,610
; FILING DATE: 07-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 31 973.8
; FILING DATE: 08-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 195 03 952.1
; FILING DATE: 07-FEB-1995
; PRIOR APPLICATION DATA: PCT/EP95/03175
; FILING DATE: 10-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bertram, Richard J.
; REGISTRATION NUMBER: 39,105
; REFERENCE/DOCKET NUMBER: P1614-7007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9318 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
US-08-793-610-6

Alignment Scores:

Pred. No.: 0 Length: 9318
Score: 5850.50 Matches: 1158
Percent Similarity: 95.00% Conservative: 59
Best Local Similarity: 90.40% Mismatches: 61
Query Match: 90.47% Indels: 3
DB: 2 Gaps: 3
US-09-672-725c-4 (1-1280) x US-08-793-610-6 (1-9318)
QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 1776 ATGATCTTGAAGGGGACCCCAATGGAGAGCAAGAAAGAAAGAACTTTTAAACTGAAC 1835
QY 20 LysLysSerLysLysGlyLysLysLysLysProThrValSerThrPheAlaMetPhe 39
Db 1836 AATAAAAGTGAAGAGATAAGAGAAACCAACTGTCAGTGATTTTCAATGTTT 1895
QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 1896 CGTATTCAAATTTGGCTTGACAGTTGTATATGTTGGTGGGAACTTTGGTGCCATCATC 1955
QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
Db 1956 CATGGGGCTGGACTTCCTCTCATGATGCTGTTTGGAGAAATGACAGATATCTTGA 2015
QY 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
Db 2016 AATGCAGGA---AATTTAGAAGATCTGATGCAACATCACTAATAGAAATGATATCAAT 2072
QY 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 2073 GATACAGGGTTCCTCATGAAT--CTGAGGAGACATGACCAGGTATGCTATATTATAC 2129
QY 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
Db 2130 AGTGAATTTGGTGTGGGTGCTGCTGCTTACATTCAGGTTTCATTTTGGTGCTG 2189
QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
Db 2190 GCAGCTGGAAGACAAATACAAAATTAGAAAACAGTTTTTTCATGCTATTAATGGACAG 2249
QY 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
Db 2250 GAGATAGCTGGTGTGATGTCACGATGTTGGGAGCTTAACACCCGACTTACAGATGAT 2309
QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
Db 2310 GTCTCTAAGATTAAAGAGTTATTGGTGACAAAATTGGAATGTTCTTTCAGTCAATGGCA 2369
QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTyrLysLeuThrLeuValIle 219
Db 2370 ACATTTTTCACCTGGGTTTATAGTAGGATTTACACGCTGGTGGAGCTAACCCCTTGTGATT 2429
QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
Db 2430 TTGGCCATCAGTCTGTTCTTGGACTGTCAGCTGCTGCTGGGCAAGATACATCTTCA 2489
QY 240 PheThrAspLysLysLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
Db 2490 TTTACTGATAAAGAACTCTTACGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGCTCTTG 2549
QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysLysLysLysLysLysLys 279
Db 2550 GCAGCAATTAGAATCTGTGATTGCTTTGGAGGACAAAAGAAAGAACTTTGAAAGGTACAAC 2609
QY 280 LysAsnLeuGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
Db 2610 AAAAATTTAGAAGAGCTAAAAGAAATTGGGATAAAGAAAGCTTATACAGCCAATTTCT 2669
QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
Db 2670 ATAGGTGCTGCTTCTGCTGATCATGATGCTTATGCTGCTGCTGCTGCTGCTGCTGCTG 2729

QY	320	SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal	339
Db	2730	ACCTTGGTCCCTCAGGGGAGATATCTATTGGACAAAGTACTCACTGTATCTTTCTGTA	2789
QY	340	LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla	359
Db	2790	TTAATTTGGGGCTTTTGTGTGGACAGCATCTCCAAGCATTTGAAGCATTTGCAANTGCA	2849
QY	360	ArgGlyAlaLaTyrgluIlePheLeuLysIleIleAspAsnLysProSerIleAspSerTyr	379
Db	2850	AGAGGAGCAGCTTATGAAATCTTCAAGATAATTGATAATAAGCCAAGTATTGCACAGCTAT	2909
QY	380	SerLysSerGlyHisLysLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis	399
Db	2910	TCGAAGAGTGGGCACAAACACAGATAATATTAGGGAAATTTGGAATTCAGAAATGTTCCAC	2969
QY	400	PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln	419
Db	2970	TTCAAGTACCACATCTCGAAAAGAGTTAGATCTTTGAAGGGCTCGAACCTCAAGGTCCAG	3029
QY	420	SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln	439
Db	3030	AGTGGGCAGACGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGACACAACTCCAG	3089
QY	440	LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle	459
Db	3090	CTGATGCAGAGGCTCTATGACCCACAGAGGGAGTGGTCACTGTTGTATGGACAGATATT	3149
QY	460	ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal	479
Db	3150	AGGACCATAAATGTAGAGTTCTTACGGGAAATCAATGGTGTGGTGAAGCACTGTA	3209
QY	480	LeuPheLaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp	499
Db	3210	TTGTTTCCACCACAGTAGCTGAAAACATTCGCTATGCGCTGAGAAATGTCACCATGGAT	3269
QY	500	GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn	519
Db	3270	GAGATTGAGAAAGCTGTCAAGGAAGCAATGCTATGACTTTATCATGAAATGCCTCAT	3329
QY	520	LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg	539
Db	3330	AAATTTGACACCTGGTTGGACAGAGAGGGGCCCAAGTTGAGTGTGGCAGAGCAGAGG	3389
QY	540	IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr	559
Db	3390	ATCGCCATTACAGTCCCTGGTTGCAACCCCAAGATCCTCCTGCTGGATGAGGCCAG	3449
QY	560	SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys	579
Db	3450	TCAGCCTTGGACACAGAAAGCAAGCAGTGTTCAGGTGGCTCTGGATAAGGCCAGAAA	3509
QY	580	GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle	599
Db	3510	GGTCGGACCAACCATGTGATACCTCATCGTTTGTCTACAGTTCTGATGTGCTACGTCATC	3569
QY	600	AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu	619
Db	3570	GCTGGTTTCGATGAGGAGTCATTTGCGAGAAAGAAATCATGATGACTCATGAAAGAG	3629
QY	620	LysGlyIleTyPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu	639
Db	3630	AAAGGCATTTACTTCAAACATTGTCACATGCAGACACAGAGAAATGAAGTTGAATTAGAA	3689
QY	640	AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer	659
Db	3690	AATGCAGCTGATGATCCAAAAGTGAATTTGATGCCCTTGGAAATGTCTTCAATGATCCA	3749
QY	660	GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln	679
Db	3750	AGATCCAGTCTAATAAGAAAAGATCAACTCGTAGGAGTGTCCGTGGATGCACAAAGCCCA	3809
QY	680	AspArgLysLeuGlyThrLysGluAspLeuAsnValProSerValSerPheTyr	699

Db	3810	GACAGAAAGCTTAGTACCAAGAGCGTCTGGATGAAGATATATACCTCCAGTTTCCCTTTGG	3869
QY	700	ArglleLeuLysLeuAsnSerThrGluTrpProTyrPheValValcIlyIlePheCysAla	719
Db	3870	AGGATATAGAACTAAATTTAACTGAATGSCCTTATTTTGTGTGGTCTATTTTGTGCC	3929
QY	720	IlelleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArglleIleIle	739
Db	3930	ATTATAATGAGCGCTGCACACAGCATTTGCAATAATATTTTCAAAAGATTATAGGGTT	3989
QY	740	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	759
Db	3990	TTTCAAGAAATTGATGATCTCTGAACACAAACGACAGAAATAGTAACTGTTTTCATATTG	4049
QY	760	PheLeuValLeuGlyIleIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	779
Db	4050	TTTCTAGCCCTTGGAATATTTCTTTTATATCAATTTTCTCTCAGGGTTTCACATTTGGC	4109
QY	780	LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg	799
Db	4110	AAAGCTGGAGAGATCTCTACCAAGCGCTCCGATACATGGTTTCCGATCCATGCTCAGA	4169
QY	800	GlnAspValSerTrpPheAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	819
Db	4170	CAGGATGTGATTTGGTTGATGACCTAAAAACACCCTGGAGCATTTGACTACCAAGCTC	4229
QY	820	AlaAsnAspAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	839
Db	4230	CCCAATGATGCTCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTGCTGTAAATTAACCAG	4289
QY	840	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr	859
Db	4290	AAATAGCAAAATCTGGGACAGAAATAATTATATCTTCATCTATGTTGGCAACTPAACA	4349
QY	860	LeuLeuLeuAlaIleValProIleIleIleAlaIleGlyValValGluMetLysMet	879
Db	4350	CTGTTACTCTTAGCAATTGTACCATCAATPGCAATAGCAGGAGTGTGTAAATGAAATG	4409
QY	880	LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr	899
Db	4410	TTGTCTGGCAACACTGAAAGATAAGAAAGAACTAGAAAGTGCTGGGAGATGCTACT	4469
QY	900	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr	919
Db	4470	GAGCAATAGAAACTTCGACACCGTTGTTCTTTTGACTCAGGACAGAAAGTTGAACAT	4529
QY	920	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	939
Db	4530	ATGTATGCTCAGAGTTTGCGGTACCATACAGAAACTCTTTGAGGAAGACACACTCTT	4589
QY	940	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	959
Db	4590	GGAAATTACATTTTCTCTCACCCAGGCAATGATGATTTTCTATGTGGATGTTTCCGG	4649
QY	960	PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe	979
Db	4650	TTTGGAGCTACTTGTGGCACAATAAATCAATGAGCTTTTGAGGATGTTCTGTAGTATTT	4709
QY	980	SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr	999
Db	4710	TCAGCTGTGTCTTTGGTGCCATGGCCGTGGGGCAAGTCAGTTCAITTTGCTCTGACTAT	4769
QY	1000	AlaIleAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1019
Db	4770	GCCAAAGCCAAATATCAGACGCCACATCATATGATCATTTGAAAAAACCCTTTGATT	4829
QY	1020	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1039
Db	4830	GACAGCTACAGACCGAGGCGCTAATGCCGNACACATTTGGAAGAAATGTACACTTGGT	4889
QY	1040	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1059

Db 4890 GAAGTTGTATTCACTATCCACCACCGGACATCCAGTCTTCAGGACTGAGCCTG 4949
Qy 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
Db 4950 GAGGTGAAGAAGCCGACAGCTGGCTCTGGTGGCGAGCAGTGGCTGTGGGAAGACACA 5009
Qy 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuLeuLeuAspGly 1099
Db 5010 GTGGTCAGCTCTGGAGCGGTTCTACGACCCCTTGGCAGGAAAGTGTGCTTATGGC 5069
Qy 1100 LysGluLeuLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyLysValSerGln 1119
Db 5070 AAGAATAAAGCGACTGAATGTTCACTGGCTCCGAGCACACTGGGCATCGTGTCCAG 5129
Qy 1120 GluProLeuLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
Db 5130 GAGCCATCTCTGTTGACTGCAGCATTTGCTGAGAACATTTGCCATATGGAGACAAAGCCGG 5189
Qy 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1159
Db 5190 GTGGTGTACAGGAAGAGATCGTGAGGGCGAGCAAGAGGCCAACATACATGCTTCATC 5249
Qy 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGln 1179
Db 5250 GAGTCACTGCCTAATAATATAGCACTAAAGTAGGAGACAAAGAACTCAGCTCTCTGGT 5309
Qy 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
Db 5310 GGCCAGAAACAGCATTTGCCATAGCTGCTGGCCCTTTGTAGACAGGCTCATATTTTGCCT 5369
Qy 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
Db 5370 TTGATGAGCCAGCTCAGCTCGGTGATACAGAAAGTGAAAGCTGTCCCAAGAGCCCTG 5429
Qy 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 5430 GACAAAGCCAGAAAGGCCAGCTGATTTGTATTGCTCACCGCTGTCCACCATCCAG 5489
Qy 1240 AsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
Db 5490 ATCGAGCTTAATATGTGGTGTTCAGATGGCAGATCAAGGAGCATGGCAGCATCAG 5549
Qy 1260 GlnLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 5550 CAGCTGTGGCACAAAGGATCTATTTTCAATGTGCTAGTGTCCAGGCTGGAACAAAG 5609
Qy 1280 Arg 1280
Db 5610 CGC 5612

RESULT 4
5206352-3
; Patent No. 5206352
; APPLICANT: Roninson, Igor B.; Pastan Ira H.; Gottesman,
; Michael M.
; TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
; SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/622,836
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 892,575
; FILING DATE: 01-AUG-1986
; APPLICATION NUMBER: 845,610
; FILING DATE: 28-MAR-1986
; SEQ ID NO: 3: 4669
; LENGTH: 5847.50
5206352-3

Alignment Scores:
Pred. No.: 0 Length: 4669
Score: 5847.50 Matches: 1157

Percent Similarity: 95.00% Conservative: 60
Best Local Similarity: 90.32% Mismatches: 61
Query Match: 90.42% Indels: 3
DB: Gaps: 3
US-09-672-725c-4 (1-1280) x 5206352-3 (1-4669)
Qy 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
Db 425 ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAAAGAAAGAACTTTTTAAACTCAAC 484
Qy 20 LysLysSerLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
Db 485 AATAAAGTGAAGAAGTAAAGAAAGAAACCACTGTCAGTGTATTTTCAATGTTT 544
Qy 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIle 59
Db 545 CGCTATTCAAATTTGGCTTGACAAGTTGTATATGGTGGGAACTTTGGCTGCCATCATC 604
Qy 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
Db 605 CATGGGCTGGACTTCCTCTCATGATGCTGTTTGGAGAAATGACAGATATCTTTGCA 664
Qy 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn 99
Db 665 AATGCAGGA---AATTTAGAAAGATCTGATGTCAAACACTACTAATAGAAAGTATCAAT 721
Qy 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
Db 722 GATACAGGGTCTTCATGNAAT---CTGGAGGAAGACATGACAGGTATGCTATATTATAC 778
Qy 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
Db 779 AGTGAATTTGGTGTGGGGTGTGCTGCTTACATTCAGGTTTCATTTTGGTGGCTG 838
Qy 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
Db 839 GCAGCTGGGAAGCAAAATACACAAATTAGAAACAGTTTTTTCATGCTATAATGCGACAG 898
Qy 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
Db 899 GAGATAGGCTGTTTGTATGTGCACGATGTGGGGAGCTTAACACCCGACTTACAGATGAT 958
Qy 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
Db 959 GTCTCTAAAGATTAACTAAGTTATTTGTCACAAATTTGNAATGTTCTTTTCAGTCAATGCA 1018
Qy 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
Db 1019 ACATTTTTCACCTGGGTTTATAGTAGGATTTACACGCTGGTTGGAAGCTAACCCCTGTGATT 1078
Qy 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
Db 1079 TTGGCCATCAGTCTCTTCTTGGACTGTGCTGCTGCTGGGCAAGATATCTTCA 1138
Qy 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu 259
Db 1139 TTTACTGATAAAGAACTCTTACGCTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTCTTG 1198
Qy 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
Db 1199 GCAGCAATTAGAACTGTGATTGCAATTTGGAGGACAAAGAAAGAACTTGAAGGTACAAC 1258
Qy 280 LysAsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
Db 1259 AAAAATTTAGAAAGAACTAAAGAAATTTGGATAAAGAAAGCTATTACAGCCCAATTTCT 1318
Qy 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
Db 1319 ATAGTCTGCTTCTTCTGCTGATCTATGATCTTATGCTCTTGGCTTCTGGTATGGAGCC 1378
Qy 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
:::|||||

Db 1379 ACCTTGGTCTCTCAGGGGAATATTCTATTGGACAAGTACTCACTGATTTCTTGTA 1438
QY 340 LeuIleGlyAlaPheSerIleGlyClnAlaSerProSerIleGluAlaPheAlaAsnAla 359
Db 1439 TTAATTTGGGCTTTTAGTGTGGACAGGCATCTCCAAGCATTTGAAGCATTTGCAAAATGCA 1498
QY 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
Db 1499 AGAGGAGCAGCTTATGAATCTTCAAGATANTTCAATATAGCCAATATGACAGCTAT 1558
QY 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
Db 1559 TCGAAGATGGGCACAAACCAAGATTAATTAAGGGAATTTGGAATTCAGAAATGTTCCAC 1618
QY 400 PheSerTyrProSerArgLysGluValLysIleLysValLysIleLysGlyCysGlyLysSerThrThrValGln 439
Db 1619 TTCAGTTACCCATCTCGAAGAAGAGTTAAGATCTTGAAGGGCTTGAACCTTGAAGGTGCAG 1678
QY 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
Db 1679 AGTGGGCAGACGGTGGCCCTGGTTGGAAACAGTGGCTGTGGGAAGAGCACACAGTCCAG 1738
QY 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
Db 1739 CTGATGCAGAGCTCTATGACCCACAGAGGGGATGGTCAGTGTTCATGACAGGATATT 1798
QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
Db 1799 AGGACCATAAATGTAAGGTTTCTACGGGAATCATTTGGTGTGTGAGTCAGAACCTGTA 1858
QY 480 LeuPheAlaThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
Db 1859 TTGTTTGGCCACCAAGATAGCTGAAACATTCGCTATGGCGGTGAAATGTCACCATGGAT 1918
QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
Db 1919 GAGATTGAGAAAGCTGTCAGAGAGGCATATGCTTATGACTTATCATGAATGCTCAT 1978
QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
Db 1979 AAATTTTGACACCTGGTTGGAGAGAGGGGCCAGTTGAGTGGTGGCAGAGCAGAGG 2038
QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
Db 2039 ATGCCATTGCAGTGCCTGGTTCGCAACCCCAAGATCCTCTGCTGGATGAGGCCAGG 2098
QY 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
Db 2099 TCAGCCTTGACACAGAAAGCGAAGCAGTGGTTTCAGTGGCTCTGGATGAAGCCAGAAA 2158
QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
Db 2159 GGTGGACCAACCATTTGTAGTATCATCTGTTGTCTACAGTTCGTAATGCTGACGTGATC 2218
QY 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
Db 2219 GCTGGTTTCGATGATGGAGTCATTTGGAGAAAGGAATCATGATGAATCATGAAAGAG 2278
QY 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
Db 2279 AAAGGCAATTTACTTCAAACTTGTCAATGTCACATGTCACAGCAGGAAATGAAGTTAGAA 2338
QY 640 AsnAlaThrGlyLysSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
Db 2339 AATGCACTGATCAATCCAAAAGTGAATTTGATGCTTGGAAATGCTTCAAAATGATTCA 2398
QY 660 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
Db 2399 AGATCCAGTCTAATGAAGAAAGATCAACTCTGAGGAGTGTCCGTGGATCATCAAGCCCAA 2458
QY 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrp 699
Db 2459 GACAGAAAGCTTAGTACCAAAAGAGCCTCTGGATGAAGATATACCTCCAGTTCTCTTTGG 2518

QY 700 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719
Db 2519 AGGATTATGAAGCTAAATTAACCTGAATGGCTTATTTTGTGTGTGTATTTTGTGCC 2578
QY 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
Db 2579 ATTATAATGAGGCGCTCAACACGACATTTGCAATAATATTTCAAAAGATATAGGGTT 2638
QY 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
Db 2639 TTTACAAAGAAATGATGATCTCTCAAAACAAACGACAGAAATAGTAACTTTTTCACATTG 2698
QY 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
Db 2699 TTTCTAGCCCTTGAATATTTCTTTTATTACATTTTCTTCCAGGGTTTTCACATTGGC 2758
QY 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
Db 2759 AAAGCTGGAGAGATCCCTCACCAAGCGCTCCGATACATGGTTTTCGATCCATGCTCAGA 2818
QY 800 GlnAspValSerThrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
Db 2819 CAGGATGTGAGTGTGTTGATGACCTTAAACACACCACCTGGACCATGACTACCAAGCTC 2878
QY 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
Db 2879 GCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTTCCAGGCTTCTGCTTAATACCCAG 2938
QY 840 AsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThr 859
Db 2939 AATATACCAATCTGGGACAGAAATATATATCTCTCATCTATGTTGGCAACTAACA 2998
QY 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
Db 2999 CTGTTACTCTTAGCAATTTGACCATCATTTGCAATAGCAGGAGTTGTTGCAATGAATG 3058
QY 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
Db 3059 TTGTCTGGACAGACCTGAAAGATAAGAAAGAACTAGAAGGTGCTGGGAAGATCCTACT 3118
QY 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
Db 3119 GAAGCAATAGAAAACCTCCGAACCGTGTGTTCTTTGACTCAGAGCAGAGTTTGAACAT 3178
QY 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
Db 3179 ATGATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTGAGGAAAGCACACATCTT 3238
QY 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
Db 3239 GGAATACATTTCTTCCCTCACCCAGGCATGATGATTTTCTATGCTGGATGTTCCGG 3298
QY 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 979
Db 3299 TTTGGAGCCTACTTGTGGTCACATAAATCATAGCTTTGAGGATGTTCTGTTAGTATT 3358
QY 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
Db 3359 TCAGCTGTGTCTTGTGGTCCCATGGCCGTGGGGCAAGTCAGTTTCATTTGCTCTGACTAT 3418
QY 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
Db 3419 GCCAAGCCAAATATCAGCAGCCCATCATCATGATCATGATGATGATGATGATGATGAT 3478
QY 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
Db 3479 GACAGCTACAGCACGGAAGGCTTAATGCCGAACATTTGGAAGGAATGTCACATTTGGT 3538
QY 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
Db 3539 GAAGTGTATTCACTATCCACCCGACCGACATCCAGTGTCTTCCAGGACTGAGCCTG 3598

us-09-672-725c-4.rni

/	CLASSIFICATION:	435
/	ATTORNEY/AGENT INFORMATION:	
/	NAME:	No. 599408nan, Kevin E
/	REGISTRATION NUMBER:	35,303
/	REFERENCE/DOCKET NUMBER:	95,1121
/	TELECOMMUNICATION INFORMATION:	
/	TELEPHONE:	312-913-0001
/	TELEFAX:	312-913-9808
/	INFORMATION FOR SEQ ID NO: 1:	
/	SEQUENCE CHARACTERISTICS:	
/	LENGTH:	4669 base pairs
/	TYPE:	nucleic acid
/	STRANDEDNESS:	single
/	TOPOLOGY:	linear
/	MOLECULE TYPE:	cDNA
/	FEATURE:	
/	NAME/KEY:	5'UTR
/	LOCATION:	1..424
/	FEATURE:	
/	NAME/KEY:	CDS
/	LOCATION:	425...4264
/	FEATURE:	
/	NAME/KEY:	3'UTR
/	LOCATION:	4265..4669
/	US-08-752-447-1	
/	Alignment Scores:	
/	Pred. No.:	0
/	Score:	5823.50
/	Percent Similarity:	94.69%
/	Best Local Similarity:	89.85%
/	Query Match:	90.05%
/	IDnls:	3
/	Gaps:	3
/	DB:	
/	US-09-672-725C-4 (1-1280) x US-08-752-447-1 (1-4669)	
QY	1 MetaspProgluglyarglysGlySerAla--GluLysAsnPheTrpLysMetGly	19
Db	425 ATGATCTTGAAGGGGCCCAATGAGGAGGAAAGAAGAACACTTTTAAACGTGAAC	484
QY	20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe	39
Db	485 AATAAAGTCGAAAAGATAGAAGGAAAAAGAAACCACCTGCAGTGCTATTTTCAATGTTT	544
QY	40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGluThrMetAlaAlaIleile	59
Db	545 CGCTATTCAAATTCGCTTGACAAGTTGTATATGGTGGGGAACTTTGGCTGCCATCATC	604
QY	60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThraspSerPheAla	79
Db	605 CATGGGGCTGGACTTCCTCTCATGATGCTGGTGGTGGAGAAATCACACATATCTTGCA	664
QY	80 AsnAlaGlylleSerArgAsnLysThrPheProValilleAsnGlnSerillethrAsn	99
Db	665 AATCAGGA---AATTAGAAGATCTGATCTCAACATCACTAATAAGAAGTGAATCAAT	721
QY	100 AsnThrGlnHisPheIleAsnHisLeuGluGluGluMetThrThrTyrAlatyrTyrTyr	119
Db	722 GATCAGGGTTCCTTCATCAAT--CTGAGAGAGACATGCCAGGTATGCCTATTATTAC	778
QY	120 SerGlylleGlyAlaGlyValLeuValAlaAlaTyrilleGlnValSerPheTrpCysLeu	139
Db	779 AGTGAATTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	838
QY	140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln	159
Db	839 GCAGCTGGAAGACAAATACACAAAATTAGAAAACAGTTTTTTCATGCTATTATCGACAC	898
QY	160 GluileGlyTTPheAspValHisAspValcglyLuleuAsnThrArgLeuThrAspAsp	179
Db	899 GAGATAGGCTGGTTGATGTCCAGCATGTTGGGAGCTTAACACCGGACTTACAGATGAT	958
QY	180 ValSerLysIleAsnGluGlylleGlyAspLysValGlyMetPheGlnSerIleAla	199

QY	920	MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe	939
Db	3179	ATGTATGCTCAGAGTTTGCAGGTACCATACAGAAATCTTTTGAGGAAGCACACATCTTT	3238
QY	940	GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg	959
Db	3239	GGAATTACATTTTCCCTACCCAGGCAATGATATTTTCTCTGCTGGAATGTTCCGG	3298
QY	960	PheGlyValTyrLeuValAlaAsnGluPheMetAsnPhenGlnAspValLeuLeuValPhe	979
Db	3299	TTTGGAGCCTACTTGGTGGCACAATAACTCATGAGCTTTGAGGATGTTCTGTTAGTATT	3358
QY	980	SerAlaIleValPheGlyAlaMetAlaValGlnValSerSerPheAlaProAspTyr	999
Db	3359	TCAGCTGTTGCTTCTTGTGTCATGGCGGTGGGCAAGTCAGTTCAATTTGCTCTGACTAT	3418
QY	1000	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle	1019
Db	3419	GCCAAAGCCCAATATCAGCAGGCCACATCATCATCATTTGAAAAAACCCTTTGATT	3478
QY	1020	AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1039
Db	3479	GACAGCTACAGCAGCGAAGCCCTAATGCCGACACATTTGGAAGAAATGTCAATTTGTT	3538
QY	1040	GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1059
Db	3539	GAAGTTGTATTCACATATCCACCCGACCGGACATCCAGTGTCTCAGGAGTCAGCCCTG	3598
QY	1060	GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr	1079
Db	3599	GAGGTGAAGAGGCGCAGACGCTGGCTCTGTGGCAGCAGTGGCTGTGGAAAGACACA	3658
QY	1080	ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly	1099
Db	3659	GTGGTCCAGCTCTCTGGACGCGTTCTACGACCCCTTGGCAGGAAAGTGTGCTTGATGGC	3718
QY	1100	LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln	1119
Db	3719	AAAGAAATAAGCCACTGAATGTTCAGTGGCTCGGACACACCTGGGCAATGCTTCCAG	3778
QY	1120	GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg	1139
Db	3779	GAGCCCATCTCTGTGTGCTGAGCATTTGCTGAGAACATTCCTATGGAGACAACGCCGG	3838
QY	1140	ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle	1159
Db	3839	GTGGTGTCCAGGAGAGAGATCGTGAGGCGCAGCAAGAGGCGCCACATACATGCTTCATC	3898
QY	1160	GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly	1179
Db	3899	GAGTCACCTGCCTAATAATATAGCACATAAGTAGGAGACAAAGAACTCAGCTCTCTGTT	3958
QY	1180	GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu	1199
Db	3959	GGCCAGAAACACCATTCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	4018
QY	1200	LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu	1219
Db	4019	TGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGAGTTGTCCAAAGAGCCCTG	4078
QY	1220	AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln	1239
Db	4079	GACAAAGCCAGAGAGGCGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	4138
QY	1240	AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln	1259
Db	4139	AATGCAGACTTAATAGTGGTGTTCAGAGTGCAGAGTCAAGGAGCATGGCCACCATCAG	4198
QY	1260	GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys	1279
Db	4199	CAGCTGCTGGCACAGAAAGGCATCTATTTTCAATGGTCTGCTGCTGCTGCTGCTGCTG	4258

QY	1280	Arg	1280
Db	4259	CGC	4261
RESULT 6			
US-08-784-649A-1			
; Sequence 1, Application US/08784649A			
; Patent No. 5830697			
; GENERAL INFORMATION:			
; APPLICANT: Sikic, Branimir I			
; APPLICANT: Chen, Gang			
; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO			
; TITLE OF INVENTION: CYCLOSPORIN MODULATION			
; NUMBER OF SEQUENCES: 5			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Fish & Richardson			
; STREET: 2200 Sand Hill Road			
; CITY: Menlo Park			
; STATE: CA			
; COUNTRY: USA			
; ZIP: 94025			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.25			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08784,649A			
; FILING DATE:			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Sherwood, Pamela J			
; REGISTRATION NUMBER: Reg. No. 5830697 36,677			
; REFERENCE/DOCKET NUMBER: 06037/007001			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 415-322-5070			
; TELEFAX: 415-854-0875			
; INFORMATION FOR SEQ ID NO: 1:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 4264 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: cDNA			
US-08-784-649A-1			
Alignment Scores:			
Pred. No.:	0	Length:	4264
Score:	5809.00	Matches:	1157
Percent Similarity:	94.93%	Conservative:	59
Best Local Similarity:	90.32%	Mismatches:	61
Query Match:	89.83%	Indels:	5
DB:	2	Gaps:	4
US-09-672-725C-4 (1-1280) x US-08-784-649A-1 (1-4264)			
QY	1	MetAspProGluGlyGlyArgLysGlySerAla--GluLysAsnPhetPrLysMetGly	19
Db	143	ATGGATCTTGAAGGGGACCGCAATGGAGGAGCAAGAAAGAACTTTTAACTGAAC	202
QY	20	LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe	39
Db	203	AATAAAGTGAAAGATGAAGAGGAAAGAAACCACTGTCTAGTGTATTTCAATGTTT	262
QY	40	ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaIleIle	59
Db	263	CGCTATTCAAAATG-CTTGACAAAGTTGATGTTGGTGGGAACTTTGGCTGCCATCATC	321
QY	60	HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla	79
Db	322	CATGGGGCTGGACTTCTCTCATGATGCTGTTGGAGAAATGACAGATATCTTTGCA	381
QY	80	AsnAlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsn	99

QY 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
 DB 2593 GCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGTTCAGGCTTGCCTGTAATACCCAG 2652
 QY 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 859
 DB 2653 AATATAGCAAACTCTGGGACAGGAATAATATATCTCTCATCTATGTTGGCAACTAACA 2712
 QY 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetIleYMet 879
 DB 2713 CTGTTACTCTTAGCAATTTGACCATTTGACCATTTGCAATGAGGAGTTGTTGAAATGAAATG 2772
 QY 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
 DB 2773 TTTCTGGAAGAGCTGAAAGATAGAAGAACTAGAAGGTCTGGGAGATCGCTACT 2832
 QY 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
 DB 2833 GAAGCAATGAAAACTTCGGAACCGTTGTTCTTCTGACTCAGGAGCAGAAAGTTTGAACAT 2892
 QY 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
 DB 2893 ATCTATGCTCAGAGTTGACAGTACCATACAGAACTCTTTGAGGAACACACATCTTT 2952
 QY 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
 DB 2953 GGAATACATTTTCCCTCACCCAGGCAATGATGATATTTTCTATGCTGGATGTTCCGG 3012
 QY 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe 979
 DB 3013 TTTGGAGCCCTGTTGGTGACATACTCATGAGCTTTGAGGATGTTCTGTAGTATTT 3072
 QY 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerPheAlaProAspTyr 999
 DB 3073 TCAGCTGTTGCTGTTGTCATGGCGGCGGCAAGTCACTTCATTTGCTCTGACTAT 3132
 QY 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
 DB 3133 GCCAAAGCCAAATATACAGCAGCCCATCATCATCATGATCATGATAAAGCCCTTTGATT 3192
 QY 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
 DB 3193 GACAGCTACAGCAGGAGGAGCCCTAATGCCGACACATTTGGAAGAAATGTCATATTTGGT 3252
 QY 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
 DB 3253 GAAGTTGTATTCATATATCCACCCGACCGGACATCCCATGCTTCAGGAGCTGAGCGTG 3312
 QY 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
 DB 3313 GAGGTGAAGAAGGCGCCAGCGCTGCTGCTGGCAGCAGTGGCTGTGGGAAGAGCACA 3372
 QY 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
 DB 3373 GTGTCCAGCTCTCGGAGCGGCTTACGACCCCTTGGCAGGAAAGTCTGCTGTATGGC 3432
 QY 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
 DB 3433 AAAGAATAAAGCAGCTGAATGTTCACTGGCTCGGACACACCTGGGCATCGTGTCCAG 3492
 QY 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
 DB 3493 GAGCCCATCTCTGTTGATGCTGAGCATTTGCTGAGAACATTTGCTATGAGAACACACCGCG 3552
 QY 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1159
 DB 3553 GTGTGTGCACAGGAAGAGATTTGAGGCGCAGCAAGGCGCCACATACATGCTTCATC 3612
 QY 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
 DB 3613 GAGTCACTGCCTAATAAATATAGCACTAAAGTAGGAGACAAAGAACTCAGCTCTCTGTT 3672

QY 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
 DB 3673 GCCAGAAACAAGCATTCGCTGCGCCTTTGTAGACAGCCTCATATTTTGGCTT 3732
 QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
 DB 3733 TTGATGAAGCCAGCTCAGCTCTGGATACAGAAAGTGAAGGTTGTCCAAAGAGCCCTG 3792
 QY 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
 DB 3793 GACAAAGCCAGAGAGCCGACCTGCATTTGATTGCTCAGCGCTGTCCACCATCCAG 3852
 QY 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
 DB 3853 AATCAGACTTAAATAGTGGTGTTCAGATGGCAGATCAAGGAGCATGGCAGCATCAG 3912
 QY 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
 DB 3913 CAGTCTGGCAGAGAAAGCATCTATTTTCAATGTCAGTGTCCAGGCTGGAAACAAAG 3972
 QY 1280 Arg 1280
 DB 3973 CGC 3975
 RESULT 7
 US-08-784-649A-5
 ; Sequence 5, Application US/08784649A
 ; Patent No. 5830697
 ; GENERAL INFORMATION:
 ; APPLICANT: Sikic, Branimir I
 ; APPLICANT: Chen, Gang
 ; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
 ; TITLE OF INVENTION: CYCLOSPORIN MODULATION
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 2200 Sand Hill Road
 ; CITY: Menlo Park
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/784,649A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sherwood, Pamela J
 ; REGISTRATION NUMBER: Reg.No. 5830697 36,677
 ; REFERENCE/DOCKET NUMBER: 06037/007001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-322-5070
 ; TELEFAX: 415-854-0875
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4264 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; US-08-784-649A-5
 Alignment Scores:
 Pred. No.: 0 Length: 4264
 Score: 5809.00 Matches: 1157
 Percent Similarity: 94.93% Conservatives: 59
 Best Local Similarity: 90.32% Mismatches: 61
 Query Match: 89.83% Indels: 5
 Gaps: 4

Qy	720	IleIleAsnGlyClyIleuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	739
Db	2293	ATTATATAATGGAGGCGTCGAACAGCATTTCCCAATAATATATTTCAAGAGTATTATAGGGTT	2352
Qy	740	PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu	759
Db	2353	TTTACAGAATATGATGATCCTGGAACAACACACAGATAGTAACITGTGTTTTCACATATTG	2412
Qy	760	PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly	779
Db	2413	TTTCTAGCCCTTGGAAATATTTCTTTATATACATTTTCTCTCAGGGTTTTCACATTTGCG	2472
Qy	780	LysAlaGlyCyluIleLeuThrLysArgLeuArgIleMetValPheArgSerMetLeuArg	799
Db	2473	AAAGCTGGAGAGATCTCTACCAAGCGGCTCCGATACATATGTTTCCGATCCATGCTCAGA	2532
Qy	800	GlnAspValSerIlePheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu	819
Db	2533	CAGGATGTGATTTGGTTTGTATGACCTTAAACACCACTGGAGCATGTACACAGGCTC	2592
Qy	820	AlaAsnAspAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln	839
Db	2593	GCCAAATGATCTCTCAAGTTAAAGGGCTATAGTTCACAGCTTGCTGTAATTAACCCAG	2652
Qy	840	AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleIleThrGlyThrGlnLeuThr	859
Db	2653	AAATATACAAATCTTGGGACAGAAATATATATCTTCATCTATGTTGGCAACTAACA	2712
Qy	860	LeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet	879
Db	2713	CTGTTACTCTTAGCAATTTGTACCCATCATPTGCAATPAGCAGGAGTGTGTAATGAAGATG	2772
Qy	880	LeuSerGlyGlnAlaLeuLysAspLysGlyGlnLeuGluGlyAlaGlyLysIleAlaThr	899
Db	2773	TTGTCTGGACAGCCTGGAAGATAGAAGAAGCACTAGAAGGTGCTGGGAAGATCGCTACT	2832
Qy	900	GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluThr	919
Db	2833	GAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAGTTTGAACT	2892
Qy	920	MetThrAlaGlnSerLeuGlnValProThrArgAsnSerLeuArgLysAlaHisIlePhe	939
Db	2893	ATGTATGCTCAGAGTTTTCAGGTACCATACAGAAACTCTTTGAGGAAGACACACATCTT	2952
Qy	940	GlyValSerPheSerIleThrGlnAlaMetMetThrPheSerThrAlaGlyCysPheArg	959
Db	2953	GGAAATTACATTTTCTTCCACCCAGGCAATGATGATTTTCTATGCTGGATGTTCCGG	3012
Qy	960	PheGlyAlaThrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuValPhe	979
Db	3013	TTTGGAGCCTACTTTGGTGGCACAATAAATCATAGAGCTTTGAGGATGTTCTGTTAGTATT	3072
Qy	980	SerAlaIleValPheGlyAlaMetAlaValIglyGlnValSerSerPheAlaProAspThr	999
Db	3073	TCAGCTGTGTCTTTGGTGCCATGGCCGTGGGGAAGTCAGTTTCATTTGCTCCTGACTAT	3132
Qy	1000	AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProIleIle	1019
Db	3133	GCCAAAGCCAAAATATCAGCAGGCCACATCATCATGATCATTTGAAAACACCCCTTGATT	3192
Qy	1020	AspSerThrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn	1039
Db	3193	GACAGCTACAGCAGGAGGCCCTAATGCCGGAACACATTCGAAGGAATGTCACATTTGGT	3252
Qy	1040	GluValValPheAsnThrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu	1059
Db	3253	GAAGTTGATTAACATATCCACCCGACCGACATCCCAAGTGCCTCAGGGACTGAGCGTG	3312
Qy	1060	GluValLysLysGlyGlnThrLeuAlaLeuValIglySerSerGlyCysGlyLysSerThr	1079
Db	3313	GAGGTCAAGAGGGCCAGACGCTGGCTCTGGTGGGCAGCAGTGGCTGTGGGAAGAGCAC	3372

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/332,444
;; FILING DATE: 31-OCT-1994
;; APPLICATION NUMBER: 07/887,712
;; FILING DATE: 22-MAY-1992
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4669 bases
;; TYPE: nucleic acid
;; STRANDEDNESS: singular
;; TOPOLOGY: linear
;; MOLECULE TYPE: Genomic DNA
;; DESCRIPTION: US-08-583-276-18

Alignment Scores:

Pred. No.: 0 Length: 4669
Score: 5790.50 Matches: 1147
Percent Similarity: 94.38% Conservative: 62
Best Local Similarity: 89.54% Mismatches: 69
Query Match: 89.54% Indels: 3
DB: 2 Gaps: 3

US-09-672-725c-4 (1-1280) x US-08-583-276-18 (1-4669)

QY 1 MetAspProGluGlyArgLysGlySerAla---GluLysAsnPheTrpLysMetGly 19
DB 425 ATGGATCTTGAAGGGGACCCGAATGGAGGAGCAAGAGAACACTTTTAAACATGAAC 484
QY 20 LysLysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPhe 39
DB 485 AATAAAAGTGAAG 544
QY 40 ArgTyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIlelle 59
DB 545 CGTATTCAAAATGGCTTGACAGTGTATATGGTGGGAACTTTGGCTGCCATCATC 604
QY 60 HisGlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAla 79
DB 605 CATGGGGCTGGACTTCTCTCATGATGCTGGTGTGGAGAAATGACAGATATCTTTGCA 664
QY 80 AsnAlaGlyIleSerArgAsnLysThrPheProValIlelleAsnGluSerIleThrAsn 99
DB 665 AATGCAGGA---AATTAGAGATCTGATGTCAAACATCATCAATAAGAAGTATCAAT 721
QY 100 AsnThrGlnHisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyr 119
DB 722 GATACAGGGTCTTCATGAAT---CTGGAGGAGACATGACCAGGTATGCCCTATTATTAC 778
QY 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
DB 779 AGTGGAAATTTGGTGTGGGTGCTGGTGTGCTTACATTCAGGTTTCATTTGGTGCCGTG 838
QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
DB 839 GCAGCTGGAGAGCAATAACACAAAATTAGAAAACAGTTTTTTCATGCTATAATGGACAG 898
QY 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspasp 179
DB 899 GAGATAGGCTGGTTGATGTGCAGGATGTGGGACGCTTAACACCGAGCTTACAGATGAT 958
QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAla 199
DB 959 GTCTCTAAGATTAATGAACATTATTTGGTGACAAAATTTGGAATGTTCTTTCAGTCAATGCA 1018
QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
DB 1019 ACATTTTTCACGTGGGTTTATAGTAGATTTACACGCTGGTGGAGCTTAACCCCTTGATTT 1078
QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
DB 1079 TTGGCCATCAGTCCTGTTCTGACATGTCAGCTGCTGCTGGGCAAGATACTACTCTCA 1138

QY 240 PheThrAspLysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeu 259
DB 1139 TTTACTGATAAAGAACTCTAGCGTATGCAAAAGCTGGAGCAGTAGCTGAAGAGGTCTTG 1198
QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
DB 1199 GCAGCAATTAGAAGCTGTGATTGCTTTGGAGGACAAAAGAAAGACTTGAAGGTACAAC 1258
QY 280 LysAsnLeuGluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
DB 1259 AAAAATTTAGAGAAGCTAAAGAAATTGGGATAAAGAAAGCTATTACACCAATATTCT 1318
QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
DB 1319 ATAGGTGCTGCTTCTCTGCTGATCTTATGCTCTGGCTCTGGCTCTGGTATGGGACC 1378
QY 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
DB 1379 ACCTTGGTCTCTCAGGGGAATATCTATTGGACAGTACTCAGTATTCTTTTCTGTA 1438
QY 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
DB 1439 TTAATTTGGGGCTTTAGTGTGGACAGGCATCTCCAAGCATTTGAAGCATTTGCAATGCA 1498
QY 360 ArgGlyAlaAlaTyrGluIlePheLysIlelleAspAsnLysProSerIleAspSerTyr 379
DB 1499 AGAGGAGCAGCTTATGAATCTTCAAGATAATTGATAATAAGCCAAGTATTGACAGCTAT 1558
QY 380 SerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHis 399
DB 1559 TCGAAGAGTGGCACAACACACATAATTAAGGGAATTTGGAATTCAGAATGTTTCAC 1618
QY 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLysValGln 419
DB 1619 TTCAGTTACCCATCTCGAAAAGAAAGTTAAGATCTTTGAAGGGCCTGAACCTGAAGTGCG 1678
QY 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
DB 1679 AGTGGGACAGCGTCCCTCGTTGGAAACAGTGGCTGTGGAGAGAGACACACAGTCCAG 1738
QY 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
DB 1739 CTGATGCAGAGCTCTATGACCCACAGAGGGGATGGTCAGTGTTCATGGACAGATATT 1798
QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
DB 1799 AGGACCAATAAATGTAAGGTTTCTACGGGAAATCAATTGGTGTGGTGAAGTCAAGAACCTGTA 1858
QY 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
DB 1859 TTGTTTGCACACGATAGCTGAAACATTCGCTATGGCCGTGAAATGTCACCATGGAT 1918
QY 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
DB 1919 GAGATTGAAAAGCTGTCAAGGAAGCCCAATGCCATGACTTTATCATGAAACTGCCCTCAT 1978
QY 520 LysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
DB 1979 AAATTTGACACCCCTGGTTGGAGAGAGAGGGCCCGCCAGTTGAGTGGTGGGACAGAGG 2038
QY 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
DB 2039 ATCGCATTTGACGCTGCCCTGGTTGCGAACCCCAAGATCTCTCTGCTGGATGAGGCCACG 2098
QY 560 SerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLys 579
DB 2099 TCAGCTTGGACACAGAAAGCAAGCAGTGGTTCAGGTGGCTCTGATAAGGCCAGAAAA 2158
QY 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
DB 2159 GGTGGGACCACTTGTGATAGCTCATCGTTGTCTACAGTTCGTAATGCTGAGGTGATC 2218
QY 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619

Db 2219 GCTGGTTTCGATGGAGTCAATGCTGGAGAAAGGAATCATGATCACTCATGAAAGAG 2278
 Qy 620 LysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleLeuLeuGlu 639
 Db 2279 AAAGGCAATTTACTTCAAACTTGTCACATGCACAGCAGGAGAAATGAAGTTGAATAGAA 2338
 Qy 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSer 659
 Db 2339 AATGCAGCTGATGAATCCAAAAGTGAATGTGATGCTTGGAAATGTCITCAAATGATCA 2398
 Qy 660 GlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
 Db 2399 AGNATCAGCTCTAATAAGAAAAGATCACTCTAGAGTGTCCGGTATCACAACCCCAA 2458
 Qy 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrp 699
 Db 2459 GACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTTGG 2518
 Qy 700 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719
 Db 2519 AGGATTATGAAGCTAATTTACTGAATGGCCTTATTTTGTGTTGGTATTTGTGCC 2578
 Qy 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
 Db 2579 ATTATAAATGGAGGCTGCAACAGCATTTGCAATAATATTTTCAAAGATTTATAGGGTT 2638
 Qy 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
 Db 2639 TTTTACAAGAATTGATGATCTGAAACAAACAGCAGCAATACAACTTGTTCACATATG 2698
 Qy 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
 Db 2699 TTTCTAGCCCTTGGAAATATTTCTTTTATTTACATTTTCTTCAGGGTTTCACATTTGGC 2758
 Qy 780 LysAlaGlyIleLeuLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
 Db 2759 AAAGCTGGAGAGATCTCTACCAAGCGGCTCCGATACATGTTTCCGATCCATGCTCAGA 2818
 Qy 800 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
 Db 2819 CAGGATGTAGTGGTTGTGATGACCTTAAACACACACATGGAGCATTGACTACCGGCTC 2878
 Qy 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839
 Db 2879 GCCAATCATGCTGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTCTGTAATATCCACG 2938
 Qy 840 AsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThr 859
 Db 2939 AATATAGCAATCTTGGGACAGGAATAATATATATCTTCACTATGTTGGCAACTAACA 2998
 Qy 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
 Db 2999 CTGTTACTCTTGAATTTGATACCATCATTGCAATAGCAGGAGTTGTTGAAATGAAGATG 3058
 Qy 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
 Db 3059 TTGTCTGGACAAAGCAGCTGAAAGATAAGAAAGAACTAGAAGGTCTGGGAGATCGCTACT 3118
 Qy 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
 Db 3119 GAACAAATAGAAAATCTTCCGACCGTGTGTTCTTGTGACTCAGGAGCAGAGTTTGAACAT 3178
 Qy 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
 Db 3179 ATGTATGCTCAGAGTTTGCAGGTACCATACAGAAACTCTTTTGAGGAAGCAGACATCTTT 3238
 Qy 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
 Db 3239 GGAATTTACATTTCTTCCACCGGCAATGATGATTTTCTTATGCTGATGTTTCCGG 3298
 Qy 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 979

Db 3299 TTTGGAGCCTACTGTTGGGCACATAAACTAATGAGCTTTGAGGATGTTCTGTAGTATTT 3358
 Qy 980 SerAlaIleValPheGlyAlaMetAlaValGlnValSerSerPheAlaProAspTyr 999
 Db 3359 TCAGCTGTGTTCTTTTGTGGCCATGGCGTGGCAAGTGAAGTTCAATTCCTCTGACTAT 3418
 Qy 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
 Db 3419 GCCAAAGCCAAAATATCAGCAGCCACATCATCATGATCATTTGAAAACCCCTTTGATT 3478
 Qy 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
 Db 3479 GACAGCTACAGCAGCGAAGGCTTAATCCGACACATGTTGAAGAAATGTCACATTTGGT 3538
 Qy 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
 Db 3539 GAAGTTGATTAACATATCCACCAGCCGACATCCAGTCTCTTCCAGGACTGAGCCTG 3598
 Qy 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
 Db 3599 GAGGTGAAGAGGCGCCAGACGCTGGCTCTGGTGGCCAGCAGTGGCTGTGGAGAGACACA 3658
 Qy 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
 Db 3659 GTGTCTCAGCTCTCGAGCGGTTCTACGACCCCTTGGCAGGAAAGTGTCTGTATGGC 3718
 Qy 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
 Db 3719 AAAGAAATAAAGCGACTGAATGTTTCAGTGGCTCCGAGCACACCTGGGCATCGTCCAG 3778
 Qy 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
 Db 3779 GAGCCCATCTCTTGTGACTGCAGCATTTGCTGAGAACATTTGCTATGACACACAGCCGG 3838
 Qy 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPheIle 1159
 Db 3839 GTGGTGTACAGGAAGAGATCGTGGGSCAGCAAGAGGCGCAACATACATCGCTTCATC 3898
 Qy 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
 Db 3899 GAGTCACTGCCTTAATAATATAGCACTAAAGTAGAGACAAGAACTCAGCTCTCTGGT 3958
 Qy 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
 Db 3959 GCCAAGAAACACGATTCGATAGTCTGCTGCTGCTGTTGTAGACAGCTCATATTTGCTT 4018
 Qy 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
 Db 4019 TTGGATGAAGCCACGCTCAGCTTGGATACAGAAAGTGAAGAGGTTGTCCAAAGAGCCCTG 4078
 Qy 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
 Db 4079 GACAAACCCAGAGAAGCGCCACCTGTCATTTGATTGCTCACCCTGCTCCACCATCCAG 4138
 Qy 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
 Db 4139 AATGCAGACTTAATAGTGTGTTTTCAGAAATGCGAGAGTCAAGCAGCATGGCAGCATCAG 4198
 Qy 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
 Db 4199 CAGCTGCTGGCAGAAAAGGCACTATTTTTCATGGTGCAGTGTCCAGCTTGGAAACAAG 4258
 Qy 1280 Arg 1280
 Db 4259 CGC 4261

RESULT 9
 US-09-120-513-1
 ; Sequence 1, Application US/09120513
 ; Patent No. 6025160
 ; GENERAL INFORMATION:
 ; APPLICANT: Brun, Kimberly
 ; APPLICANT: Chenery, Richard


```
Db 3611 CTGGATGAAGCGACATCAGCTCTGGATACGGAGAGTGAAGAGGTCTCCAGGAAGCGCTG 3670
QY 1220 AspLysAlaAtgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
Db 3671 GACAAAGCCAGGAAGCGACCTGCAATGTGATCGCGCACCGCTCTCCACCATCCAG 3730
QY 1240 AsnAlaAspLeuIleValIlePheGlnAsnGlyLysValIleHisGlyThrHisGln 1259
Db 3731 AACGCAGACTTGATCGTGGTATTCAGAACGCCAGGTCAGAGGACGGACCCACCAG 3790
QY 1260 GlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 3791 CAGCTGCTGGGCCAGAAAGGCATCTATTCTCGATG-----GTTCAAGCTGGAGCAAAG 3844
QY 1280 Arg 1280
Db 3845 CGC 3847

RESULT 10
US-09-450-105-1
; Sequence 1, Application US/09450105
; Patent No. 6169166
; GENERAL INFORMATION:
; APPLICANT: Kimberly Anne Brun
; APPLICANT: Richard James Chenery
; APPLICANT: Harma Ellens
; APPLICANT: John Anthony Feild
; APPLICANT: Lin Yue
; TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: ENCODING RAT MDR1b2 AND SCREENING METHODS THEREOF
; FILE REFERENCE: GP-50008-D1
; CURRENT APPLICATION NUMBER: US/09/450,105
; EARLIER FILING DATE: 1999-11-29
; EARLIER FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4233
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-450-105-1

Alignment Scores:
Pred. No.: 0 Length: 4233
Score: 5296.00 Matches: 1030
Percent Similarity: 90.63% Conservative: 131
Best Local Similarity: 80.41% Mismatches: 112
Query Match: 81.89% Indels: 8
DB: 4 Gaps: 5

US-09-672-725C-4 (1-1280) x US-09-450-105-1 (1-4233)
QY 1 MetAspProGluGlyArgLysGlySerAlaGluLysAsnPheTrpLysMetGlyLys 20
Db 26 ATGGAGTTTGAAGAGGCGCTTACGGAACAGCAGCAAGAACTCTCAAGAGATGGGCAAA 85
QY 21 LysSerLysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPheArg 40
Db 86 AAGAGTAAAGAGGAG--AAGGAGAGAAACCTGCTGTGGCATATTCGGGATGTTTCGC 142
QY 41 TyrSerAsnTrpLeuAspArgLeuTyrMetLeuValGlyThrMetAlaAlaIleIleHis 60
Db 143 TATGCAGATTGGCTTGACAAAGCTGTGATGGCTCTGGGAACCTCGCTCTCATCCAC 202
QY 61 GlyAlaAlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAsn 80
Db 203 GGAACCCCTGCTCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 262
QY 81 AlaGlyIleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsn 100
Db 263 GCA-----GAGACCCGCATTCTGGCCGAGCGGTACTTAATCAAGTGAATCAACAGT 313
```

```
QY 101 ThrGlnHisPheIle---AsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyr 119
Db 314 ACACAGACCCGTCACGACGACGACGCTCTGGAGGAGACATAGCCCATGTACGCGCTACTATTAC 373
QY 120 SerGlyIleGlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeu 139
Db 374 ACGGGCATTTGGTCCGGGTGCTCATCGTTCCTTACATCCAGGTTTTCATCTTTGGTGGCTG 433
QY 140 AlaAlaGlyArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGln 159
Db 434 GCAGCTGGGAGACAAATACACAAGATTAGCAGAAAGTTTTTCCATGCCATCATGAATCAG 493
QY 160 GluIleGlyTrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAsp 179
Db 494 GAGATAGGCTGGTGTGACGTGAATGACGCTGGGAGCTCAACACCCGGCTCACAGATGAC 553
QY 180 ValSerLysIleAsnGluGlyIleGlyAspLysValIleGlyMetPheGlnSerIleAla 199
Db 554 GTCTCCAAATTAATGACGGAATTTGGTACAAACTTTGGAATGTCTTTCAGTCCATTAACG 613
QY 200 ThrPhePheThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIle 219
Db 614 ACATTTTTCAGCCGCTTTTATAATAGGATTTATAGTGGTTGGAGCTAAACCTTTGTAAT 673
QY 220 LeuAlaIleSerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSer 239
Db 674 TTGGCGCTCAGCCCTCTTATTGGTGTTCATCTGCCATGTGGCAAGTACTGACTTCA 733
QY 240 PheThrAspLysGluLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeu 259
Db 734 TTTTACTTAATAAGAACTCCAGGCTTATGGAAGCTGGAGCAGTGGCCGAGAGTCTTFA 793
QY 260 AlaAlaIleArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsn 279
Db 794 GCAGCCATCAGAACTGTGATGCTGGTTGGAGGACAAAAGAACTTGAAGGTACAAT 853
QY 280 LysAsnLeuGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSer 299
Db 854 AAAAATTTAGAGAGCTAAAGAGTTGGCTAAAGAAAGCCATCAGGCCCAACATTTCC 913
QY 300 IleGlyAlaAlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThr 319
Db 914 ATAGGTATTGCTACCTGTTGGTCTATGGTCTTATGCATGCGATTTCTGGTATGGGACC 973
QY 320 SerLeuValLeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerVal 339
Db 974 TCTTTGGTCTCTCAAAATGAATATTCTATTGGACAAGTGTTCACCGTCTTCTCTCTATT 1033
QY 340 LeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAla 359
Db 1034 TTATTGGGACCTTTTCAGTATTGGACATTTAGCCCAACATAGAACCTTTTGCAATGCA 1093
QY 360 ArgGlyAlaAlaTyrGluIlePheLysIleIleAspAsnLysProSerIleAspSerTyr 379
Db 1094 AGAGGGGACGCTATGAAATCTTCAAGATAATTGATAATGAGCCCAAGCATTCACAGCTTC 1153
QY 380 SerLysSerGlyHisLysProAspAsnIleLysGlyLysLeuGluPheLysAsnValHis 399
Db 1154 TCAACAAAGGGGACACAAACCCAGACAGTAAATGGGAAATTTGGAATTTAAAAATGTTTAC 1213
QY 400 PheSerTyrProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGln 419
Db 1214 TTCACACTACCCATCAGAGTGAAGTTAAGATCTTTGAAGGGCCCTCAACCTGAAGGTGAAG 1273
QY 420 SerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGln 439
Db 1274 AGCGGCGACAGCGTAGCCCTGGTTGGCAACAGTGGCTGTGGGAAAGACACAACTGTCCAG 1333
QY 440 LeuMetGlnArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIle 459
Db 1334 CTGCTGCAGAGGCTCTACGACCCCATAGAGCGGAGGTTCAGTATCGAGGACAGGACATC 1393
QY 460 ArgThrIleAsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProVal 479
```

Db 1394 AGGACCATCAATGTGAGGTATCTGCGGAAATCATTTGGGGTGGTGGAGTCAGAAACCCGNG 1453
 Qy 480 LeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArgGluAsnValThrMetAsp 499
 Db 1454 CTGTTGGCCACACGATTCGCGAAACATTCGCTATGCGCGAGAAACCTCACCATGGAT 1513
 Qy 500 GluIleGluLysAlaValLysGluAlaAsnAlaTyrAspPheIleMetLysLeuProAsn 519
 Db 1514 GAGATAGAGAAAGCTGTCAAGGAAGCAATGCGCTATGACTTCATCATGAAGCTGCCCCAC 1573
 Qy 520 LysPheAspThrLeuValGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArg 539
 Db 1574 AAATTTAACACCTCGTGTGTGAGAGGGCGGCACCTGAGTGGGGACAGAAACAGAG 1633
 Qy 540 IleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThr 559
 Db 1634 ATGCCCATTCGCCGGCCCTGGTCCGCAACCCCAAGATCCTTTTGTGGATGAGGCCAG 1693
 Qy 560 SerAlaLeuAspThrGluSerGluAlaValAlaValGlnValAlaLeuAspLysAlaArgLys 579
 Db 1694 TCAGCCTTGGACACAGAAAGCGGTGTTTCAGGCCCGCTCTGGATAAGGCTAGAGAA 1753
 Qy 580 GlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIle 599
 Db 1754 GCCCGGACCAACCATGTGATAGCTACCCGCTGTCTACAGTGGCAATGCTGACGTCAT 1813
 Qy 600 AlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGlu 619
 Db 1814 GCTGGTTTGTATGTTGTTGTCATTTGGAGCAAGGAAATCATGAAGCTCATGAAGAG 1873
 Qy 620 LysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGlu 639
 Db 1874 AAGGCAATTTACTTCAAACTTGTATGACACAGACATAGAGAAATGAATTTGAACAGGA 1933
 Qy 640 AsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuMetSerProLysAspSer 659
 Db 1934 AATAAGCTTATGATCCCAAGTGCACATGGTGGCTCTGAGTGTACTTCAGAAATAATCA 1993
 Qy 660 GlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGln 679
 Db 1994 AAATCTCTCTTAAATA---AGGAGATCAATTCGCAGAAAGTATCCACAGAAAGACAC 2050
 Qy 680 AspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheThr 699
 Db 2051 GAGAGAAGACCTAGTTCGAAAGAGATGTGGATGAAGATGTCCTATGCTTCTCTTTGG 2110
 Qy 700 ArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719
 Db 2111 CAGATCCTTAAAGCTAAATATATTAGTGAATGGCCCTATTAGTTGTGGGTGACTTTGTGCT 2170
 Qy 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
 Db 2171 GTTATAAATGGGTGCATACACACAGTGTTCCTCATATGTTTTCAAAGATGTAGGGGTT 2230
 Qy 740 PheThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeu 759
 Db 2231 TTTTCAAGAGACGACGACCATGAACCAACAAACAGGAATTTGTAATTTTCCCTTCTC 2290
 Qy 760 PheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGly 779
 Db 2291 TTTCTGGTCATGGGAATGATTTCTTTGTTACGTACTTCTTTCAAGGCTTCACATTTGGC 2350
 Qy 780 LysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArg 799
 Db 2351 AAAGCTGGAGAGATCCTCACCAGCGCATCCGATACATGCTTCAATCCATCCTCGGCA 2410
 Qy 800 GlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeu 819
 Db 2411 CAGGATATAAGCTGTTTGTATGATCAACCAACCAACCACTGGCTGCTGACTTACCAGGCTC 2470
 Qy 820 AlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGln 839

Db 2471 GCTAGTGACGCTTCTAATGTTAAAGGGGCTATGGGCTCCAGGCTTGCTGTAGTTACCCAG 2530
 Qy 840 AsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTyrGlnLeuThr 859
 Db 2531 AATGTAGCAAAACCTTGGCACAGGAATTATCTTCCCTTAGTCTATGGCTGGCAGCTTACA 2590
 Qy 860 LeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMet 879
 Db 2591 CTTTACTTGTAGTAATTATACCACTCATCTCTTGGTGGAAATTATTGAATGAACACTG 2650
 Qy 880 LeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThr 899
 Db 2651 TTGCTGTGCTCAAGCCTTGAAGCAACAAGAGCTAGAGATCTCTGGGAAGATCGCTACA 2710
 Qy 900 GluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGluTyr 919
 Db 2711 GAAGCAATTTGAAACTTCCGCACCTGTCTCTTTGACTAGGAGCAGAGTTTGAACACT 2770
 Qy 920 MetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePhe 939
 Db 2771 ATGTATGCCAGAGCTTGCAGATACCATACAGAAATGCTTTGAAGAAAGCACAGCTCTT 2830
 Qy 940 GlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArg 959
 Db 2831 GGGATCACCTTCGCTTCAACCCAGGCCATGATTTATTTTCTATGCTGCTGTTTCCGG 2890
 Qy 960 PheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPhe 979
 Db 2891 TTCGGTGCTACTTGGTGGCAGCAACTCATGACGTTTGAAGATGTTATGTTGGTATTT 2950
 Qy 980 SerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAspTyr 999
 Db 2951 TCTGCTGTTGTTTGGTGGCATGGCAGCAGGAATACAGTTCATTCCCTCCCTGACTAC 3010
 Qy 1000 AlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIle 1019
 Db 3011 GCGAAGGCCAAAGTCTCAGCATCCCACTCCAGATCATCAGATCATTGAGAAATCCCGAGAT 3070
 Qy 1020 AspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsn 1039
 Db 3071 GACAGCTACAGCACGCGAGGCTTGAAGCCTAATGTTGTAAGAGAAATGTGAATTTAAT 3130
 Qy 1040 GluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSerLeu 1059
 Db 3131 GGAGTCATGTTCAACTATCCACCCGACCAACATCCCACTCCAGTCTCCAGGACTGAGCTTC 3190
 Qy 1060 GluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThr 1079
 Db 3191 GAGTGAAGAAGGGCAACGCTTCGCTGGTGGCAGCAGTGGCTGGGGAAGAGTACA 3250
 Qy 1080 ValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAspGly 1099
 Db 3251 GTGGTCCAGCTGCTCAGCGCTTCTACAAACCCCATGGCTGGCAACAGTGTCTAGATGGC 3310
 Qy 1100 LysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSerGln 1119
 Db 3311 AAAGAATAAAACAACTCAACGTCCTAGTGGCTCCGCGCCCACTGGGCATTTGTCGCCAG 3370
 Qy 1120 GluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
 Db 3371 GACCCCATCTGTTGCTGACATCACCGAGAACATCGCTTACGAGACAACAGCCGT 3430
 Qy 1140 ValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisPheIle 1159
 Db 3431 GTGCTGTCTCATGAGAGATCGTGAAGGCCCGCGAGGGCCCAACATCCACCACTTCATC 3490
 Qy 1160 GluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGly 1179
 Db 3491 GACTACTGCTCAGAAATACACACAGAGTGGGAGACAAAGGGACTCAGCTGTCCGGC 3550
 Qy 1180 GlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeuLeu 1199
 Db 3551 GGGCAGAGACGCGCATCGCCATCGCGCGCCCTCTGTCAGACAGCCTCACACTTACTTT 3610

QY 1200 LeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAlaLeu 1219
|||||
Db 3611 CTGGATGATGAGGACATAGCTCTGATACGGAGAGTGAAGAGTCTCCAGGAGCGCTG 3670
QY 1220 AspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGln 1239
|||||
Db 3671 GACAAAGCAGGAAGCGCACCTGCTGATCGCGCACCGCTGTCCACCATCCAG 3730
QY 1240 AsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHisGln 1259
|||||
Db 3731 AACGAGACTGATCGTGGTGATTCAGAACGGCCAGGTCAAGGACGACGCCACCCACAG 3790
QY 1260 GlnLeuLeuAlaGlnLysGlyIleThrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
|||||
Db 3791 CAGCTGCTGCGCCAGAAAGCATCTATTCTCGATG-----GTTAGGCTGGAGCAAG 3844
QY 1280 Arg 1280
|||
Db 3845 CGC 3847
RESULT 11
US-08-461-823-1
; Sequence 1, Application US/08461823
; Patent No. 593840
; GENERAL INFORMATION:
; APPLICANT: Bhatnagar, Satish K.
; APPLICANT: George Jr., Albert L.
; APPLICANT: Nazarenko, Irina
; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OncorPharm, Inc.
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,823
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/168,621
; FILING DATE: 16-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,433
; FILING DATE: 27-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Karta, Glenn E.
; REGISTRATION NUMBER: 30,649
; REFERENCE/DOCKET NUMBER: PA-0012 CIP 2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301 527-2058
; TELEFAX: 301 208-6997
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2726 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-461-823-1

Alignment Scores:

Pred. No.: 0 Length: 2726

Score: 3615.00 Matches: 713
Percent similarity: 96.15% Conservative: 37
Best Local Similarity: 91.41% Mismatches: 30
Query Match: 55.90% Indels: 0
DB: 1 Gaps: 0
US-09-672-725C-4 (1-1280) x US-08-461-823-1 (1-2726)
QY 501 IleGluLysAlaValLysGluAlaAsnAlaThrAspPheIleMetLysLeuProAsnLys 520
|||||
Db 2 ATTGAGAAAGCTGTCAGGAAGCCCAATGCCATGACCTTTATCATGAACCTGCCTCAFAAA 61
QY 521 PheAspThrLeuValGlyGluArgGlyAlaArgLeuSerGlyGlyGlnLysGlnArgIle 540
|||||
Db 62 TTTGACACCCCTGGTGGAGAGAGAGGGCCAGTTGAGTGGTGGGAGAGCAGAGGATC 121
QY 541 AlaIleAlaArgAlaLeuValArgAsnProLysIleLeuLeuLeuAspGluAlaThrSer 560
|||||
Db 122 GCCATTGCACGTGCCCTGGTTCGCAACCCCAAGATCCTCTCTGGATGAGGCCAGTCA 181
QY 561 AlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArgLysGly 580
|||||
Db 182 GCCTTGGACACAGAAAGCGAAGCAGTGGTTCAGGTGGCTCTGGATAAGCCAGAAAAGCT 241
QY 581 ArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspValIleAla 600
|||||
Db 242 CGGACCACCATTTGATAGTACATCGTTGCTACAGTTCGTAATGCTGACGTCATCGCT 301
QY 601 GlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLysGluLys 620
|||||
Db 302 GGTTCGATGATGAGTCATTTGGAGAAAGGAAATCATGATCACTCATGAAAGAGAAA 361
QY 621 GlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleGluLeuGluAsn 640
|||||
Db 362 GGCATTTACTTCAAACTTGTCACAATGCAGACAGCAGGAATGAATGAATGAATAAAT 421
QY 641 AlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAspSerGly 660
|||||
Db 422 GCAGCTGATGAATCCAAAGTGAATTTGATGCTTGGAAATGCTTCAATGATTCAAGA 481
QY 661 SerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAsp 680
|||||
Db 482 TCCAGTCTAATAAGAAAGATCACTCGTAGGAGTGTCCTGGATCAACACCCCAAGAC 541
QY 681 ArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheTrpArg 700
|||||
Db 542 AGAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGATATACCTCCAGTTCCCTTTGGAGG 601
QY 701 IleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAlaIle 720
|||||
Db 602 ATTATGAAGCTAAATTAATACTGAATGGCCTTATTTGTTGGTGTATTTGTGCCATT 661
QY 721 IleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePhe 740
|||||
Db 662 ATAAATGGAGGCTGCAACCCAGCATTTGCAATAATTTTCAAGATTAATAGGGGTTTT 721
QY 741 ThrArgAspGluAspProGluThrLysArgGlnAsnSerAsnMetPheSerValLeuPhe 760
|||||
Db 722 ACAAGAAATTCATGATCTCTGAAACAAAACACAGAGATAGTAACTTTTTCATATTGTTT 781
QY 761 LeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLys 780
|||||
Db 782 CTAGCCCTTGGAAATTAATTTCTTTTATACATTTTCTTCAGGGTTTTCACATTTGGCAAA 841
QY 781 AlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGln 800
|||||
Db 842 GCTGAGAGATCTCTACCAAGCGGCTCCGATACATGGTTTTTCCGATCCATCTCAGACAG 901
QY 801 AspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArgLeuAla 820
|||||
Db 902 GATGTGAGTTGGTTGATGACCCCTAAAAACACCACTGGAGCATTTGACTTACCAAGCTGCC 961
QY 821 AsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsn 840

Db 2326 CTTGGTTACATGCTCATTTGGCTGAGCTTTCTCGTTCCTTGCCTGGGGTGGGTCAACTACC 2385
Qy 729 PheSerIlePheSerArgIleIleGlyIlePheThrArgAspGluAspProGluThr 748
Db 2386 CAGGCTTTTCTGTACGCCAAGCAATATACAGCGCTCTGTTACCCGAATCAATGTTTCCAC 2445
Qy 749 Lys---ArgGlnAsnSerAsnMetPheSerValLeuPheLeuValLeuValIleIleSer 767
Db 2446 AAACCTCAGGCATCAGCGCAATTTCTGCTGATCTTCTGCTGTTGGAATGCTCAA 2505
Qy 768 PheIleThrPhePheLeuGlnGlyPheThrPheGlyLysAlaGlyGluIleLeuThrLys 787
Db 2506 TTTATCAGCCTGCTCAATCAATGATACAGCATTTGCTATTGTTCCGAGAGATCATTCGC 2565
Qy 788 ArgLeuArgTyrMetValPheArgSerMetLeuArgGlnAspValSerTrpPheAspAsp 807
Db 2566 CGAGCTAGAAGTCAAGCGTTTAGATCGATTCTTCGTCAGGACATCTCAATTTTCGACAGG 2625
Qy 808 ProLysAsnThrThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaAlaGlnValLys 827
Db 2626 GAAGAGAACACGACCGGTGCTTGCAGCTCTTCTTATCAACAGAGACGAGAAATCTGTGCG 2685
Qy 828 GlyAlaIleIleSerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGly 847
Db 2686 GCGCTTAGCGGAGTACTCTCGGCACAATCATCATGACTAGCATCTACGCTTGGGGCGCG 2745
Qy 848 IleIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuLeuAlaIleValPro 867
Db 2746 ATGATCATTCGATTCGCGATCGGATCGAAGCTGGCTCTGTTTGCATATTTCTGTCGTCCA 2805
Qy 868 IleIleAlaIleAlaGlyValValGluMetLysMetLeuSerGlyGlnAlaLeuLysAsp 887
Db 2806 ATCCTTCTGGCATCGCGCTTCCCTCAGATTCTACATGCTTGCTCAATCCAGCAACGATCG 2865
Qy 888 LysLysGluLeuGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThr 907
Db 2866 AAGTCTGCTACGAAGGTCTGCGAGCTATGCTTCCGAAGCCACGTCAGCGATCCGCAC 2925
Qy 908 ValValSerLeuThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSerLeuGlnVal 927
Db 2926 GTAGCATCCTACCTCGCAACAAGATGCTGGGGCGCTTTACACGACGATCAAAAA 2985
Qy 928 ProTyrArgAsnSerLeuArgLysAlaHisIlePheGlyValSerPheSerIleThrGln 947
Db 2986 CAGGACGGAGAGTTGATCTCAGTCTGAGATCCTCCCTGCTGTATGCTGCTGTCGAC 3045
Qy 948 AlaMetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsn 967
Db 3046 GCATTGGTGTGTTTCTGCGTCTGCTTGGGCTTCTGGTATGGTGTACATCTTAGGCCAT 3105
Qy 968 GluPheMetAsnPheGlnAspValLeuLeuValPheSerAlaIleValPheGlyAlaMet 987
Db 3106 CATGAGTACACATCTTCGCTTCTTCTGCTGCTTTCTGAGATCTTTTGGTGGCGAA 3165
Qy 988 AlaValGlyGlnValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAla 1007
Db 3166 TCAGCGGGAAGTGTCTCTCTTTTCCCGGACATGAGTGAAGCAAGAAATGCGCGTCT 3225
Qy 1008 HisValIleIleIleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeu 1027
Db 3226 CAATTCAGAAACTCTTCGACAGCAAGCCCAACCATTCATCATCTGCTGGATGAGGCGAG 3285
Qy 1028 LysProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThr 1047
Db 3286 AAGTTGGAGTCTATGAAGCGGAATCGAATTCGGGACGCTCCACTTTAGGTACCCCAAG 3345
Qy 1048 ArgProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeu 1067
Db 3346 CGGCGGAGCAGCCTGTTCTTCGAGAGCTGAATTTGACGCTGAAGCGTGGACAAATAT 3405
Qy 1068 AlaLeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPhe 1087

Db 3406 GCCCTTGTGGACCCAGCTGGATGCGGTAAGAGCACTACGATTGCTGTGTTGAGCGGATTT 3465
Qy 1088 TyrAspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnVal 1107
Db 3466 TATGACGCACATGCTGGAGGGGTCTTCGTGACGGAAGGACATTACCAAACTCAATGTC 3525
Qy 1108 GlnTrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSer 1127
Db 3526 AACTCATACCCAGCTTCTCTCCCTTGTGACCAAGAACCTACTCTGTATCAGGGTACC 3585
Qy 1128 IleAlaGluAsnIleAlaTyrGlyAspAsnSerArgValValSerHisGluGluIleMet 1147
Db 3586 ATCAAGGAAAATATCTCTGAGTCGATGAAGATGACGTTTCGGAGGAGACTTTCGATT 3645
Qy 1148 GlnAlaAlaLysGluAlaAsnIleHisHisPheIleGluThrLeuProGluLysTyrAsn 1167
Db 3646 AAGGCTCAAGATGCCAATCATATGATTCGTATGTCACCTCCCTGAGGGATTTGAC 3705
Qy 1168 ThrArgValGlyAspLysGlyThrGlnLeuSerGlyGlnLysGlnArgIleAlaIle 1187
Db 3706 ACCGTCGTTGGCAGCAGGAGGATGTTGCTGTGGTGGACAAACACGCGTGTCCGCAAT 3765
Qy 1188 AlaArgAlaLeuValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeu 1207
Db 3766 GCTCGTCCCTCTCGTGACCCCAAGCTCTTCTCTGGATGAAGCCACATCTGCTCTT 3825
Qy 1208 AspThrGluSerGluLysValValGlnGluAlaLeuAspLysAlaArgGluGlyArgThr 1227
Db 3826 GACTCCGAATCTGAGAAAGTCGTACAGCTCGCTGGATGCTGCCCGCGGGGACA 3885
Qy 1228 CysIleValIleAlaHisArgSerThrIleGlnAsnAlaAspLeuIleValValPhe 1247
Db 3886 ACGATTGCTGTGCCCGCTGAGCACCACATTCAAAACGCTGATATAATTTACGTGTT 3945
Qy 1248 GlnAsnGlyLysValLysGluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIle 1267
Db 3946 GACCAGGCAAGATCGTGAAGTGAAGCCAGCACGAGTTGATTGCAAAACAAGGCGCG 4005
Qy 1268 TyrPheSerMetIleSerValGlnAlaGlyAlaLys 1279
Db 4006 TATTACGAGCTGCTCAATCTGCAGAGTCTCGGAAG 4041

RESULT 13
US-08-996-545-1
; Sequence 1, Application US/08996545
; Patent No. 5928898
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene atrd of
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Eli Lilly and Company
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,545
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872

REFERENCE/DOCKET NUMBER: X-11766
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3334
TELEFAX: 317-276-2763

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 4002 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..4002

US-09-672-725c-4

US-08-996-545-1

Alignment Scores:

Pred. No.: 1.27e-265 Length: 4002
Score: 2482.00 Matches: 538
Percent Similarity: 59.41% Conservative: 223
Best Local Similarity: 42.00% Mismatches: 458
Query Match: 38.38% Indels: 62
DB: 2 Gaps: 8

US-09-672-725c-4 (1-1280) x US-08-996-545-1 (1-4002)

QY 23 LysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPheArgTyrSer 42
DB 268 AAGACGACGCGAGGAGATCAAGTAACATCTCTCTCGGTCTCTCGCGTATGCA 327
QY 43 AsnTrpLeuAspArgLeuValGlyThrMetLeuAlaAlaIleIleHisGlyAla 62
DB 328 ACAAAGTGGATATCATCATGTAATCAGTACATCTGTCATCTGCGCATGTCGCGCGTCG 387
QY 63 AlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAlaGly 82
DB 388 ACTTCCAGAGGATAATGTTATAT-----411
QY 83 IleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsnThrGln 102
DB 412 -----CAA 414
QY 103 HisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyrSerGlyIle 122
DB 415 ATCTCGTACGACGAGTCTATGATGAATTCACCAAGACGCTAGTCTGCTATACCTC 474
QY 123 GlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeuAlaAlaGly 142
DB 475 GGTATCGCGCGAGTTGTCTACTGTCTATGTTAGTACTGTTGGCTTCTATACCGGAGAA 534
QY 143 ArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGluIleGly 162
DB 535 CACGCCACGAGAGATCCCGGAGTATACCTTGTAGTCTATCTCGCGCAGAACATGGC 594
QY 163 TrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAspValSerLys 182
DB 595 TATTTTGAATAACTCGGTGCGGGAAGTGACCCCGTATACACGCCGATACAAACCTT 654
QY 183 IleAsnGluGlyIleGlyAspLysValGlyMetPhePheIleSerIleAlaThrPhePhe 202
DB 655 ATCCAGATGGCATTTTCGGAAGAGGTGCGTCTCAGTCTTGTACTGCCCTCGCGACATTCG 714
QY 203 ThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIleLeuAlaIle 222
DB 715 ACAGCATTCATATCCGCTACGTCAAATACTGGAAGTGGCTTAATTTGCGAGCTCAACA 774
QY 223 SerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPheThrAsp 242
DB 775 ATGTGGCCCTCGTCTCACCATTGGCGGTGTTCTCAGTTTATCATCAAGTACAGCAAA 834
QY 243 LysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluValLeuAlaAlaIle 262

DB 835 AAGTCGCTTGACAGCTACGGTGCAGCGGCACCTGTCGGAAGAGGTTCATCAGCTCCATC 894
QY 263 ArgThrValIleAlaPheGlyGlyGlnLysLysGluLeuGluArgTyrAsnLysAsnLeu 282
DB 895 AGAATGCCACAGCGTTTGCACCCCAAGACAGCTTCCGAAGCAGTATGAGGTCCACTTA 954
QY 283 GluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSerIleGlyAla 302
DB 955 GACGAAGCTGAGAAATGGGGAACAAAGACAGATTGTCTATGGTGTTCATGATTGGCGCC 1014
QY 303 AlaPheLeuLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSerLeuVal 322
DB 1015 ATGTTTGGCCCTTATGTACTCGAATCAGGTCTTGGCTTCGAGTGGGTCTCGTTTCCCTG 1074
QY 323 LeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerValLeuIleGly 342
DB 1075 GTAGATGGTCAGTCGATGGGTGATATTCACAGTTCTCATGATGCCCATTTGATCGGA 1134
QY 343 AlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArgGlyAla 362
DB 1135 TCCTTCTCTTGGGGAACGTTAGTCCAAATGCTCAAGCATTTTACAAACGCTGTGGCGCGG 1194
QY 363 AlaTyrGluIlePheLysIleLeuAspAsnLysProSerIleAspSerTyrSerLysSer 382
DB 1195 GCCGCAAGATATTTGGAACGATCGATCGCCAGTCCCATTTAGTATCCATTTGCAACGAA 1254
QY 383 GlyHisLysProAspAsnLysLysGlyAsnLeuGluPheLysAsnValHisPheSerTyr 402
DB 1255 GGAAGACGCTCGACCATTTTGGAGGCCACATTTAGTTACGCAATGTCAGCATATTTAC 1314
QY 403 ProSerArgLysGluValLysIleLeuLysGlyLeuAsnLeuLysValGlnSerGlyGln 422
DB 1315 CCATCTAGACCGAGGATCCACCTCATGGAGGATGTTCTCTGTCATGCTCCGCTCGAAAA 1374
QY 423 ThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeuMetGln 442
DB 1375 ACAACGCTTAGTCGCGCCCTCTGGCTCTGGAAAAAGTACGCGTGGCTGGCTTGGTTAG 1434
QY 443 ArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArgThrIle 462
DB 1435 CGATTCTACATGCTCTTCGCGGTACGGTTCGTCGATGGCCATGACATCAAGACCTC 1494
QY 463 AsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAla 482
DB 1495 AATCTCCGCTGGCTTCGCCAACAGATCTCTTTGGTTAGCGAGGCGCTTCTTTTGGC 1554
QY 483 ThrThrIleAlaGluAsnIleArgTyrGly-----ArgGluAsnValThr 497
DB 1555 ACGACGATTTATAAGAATATTAGGCACGCTCTCATCGGCACAAAGTACGAGATGAATCC 1614
QY 498 MetAspGlu-----IleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 513
DB 1615 GAGGATAAGGTCGCGGAACATCATCAGAACGCGCAAAATGGCGAATGCTCATGACTTT 1674
QY 514 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSer 533
DB 1675 ATTACTGCTTCCCTGAGGTTATGAGACCAATGTGGCGAGCGGTGCTTCTCTCTTCA 1734
QY 534 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu 553
DB 1735 GGTGGCCAGAAACAGCGCATTCGCAATCGCCGTGCGTGTGTAGTGACCCAAAATCTCG 1794
QY 554 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValGlnValAla 573
DB 1795 CTCCTGGATGAGCTTACTTCGCGCTTGACACAAATCCGAAGCGGTGTTCAAGACGCT 1854
QY 574 LeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrVal 593
DB 1855 TTGGAGGCGGACGCTGAAGCGCGAATCTACTATTGTGATCGCTCATGCGCTTTCACCATC 1914
QY 594 ArgAsnAlaAspValIleAlaGlyPheAspGlyValIleValGluLysGlyAsnHis 613
DB 594 ArgAsnAlaAspValIleAlaGlyPheAspGlyValIleValGluLysGlyAsnHis 613

Db	1915	AAAACGGCGCACAAACATTTGGTCTCGTCAATGGCAAAATTTGCTGAACAAGGAACCTCAC	1974
Qy	614	AspGluLeuMetLysGlnLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGly	633
Db	1975	GATGAATTTGGTTGACCGCGGAGCGCTTATCGAAACTTTGTGGAGGCTCAACGTATCAAT	2034
Qy	634	AsnGluIleLuu-----LeuGluAsnAlaThrGlyGluSerLysSerGluSerAsp	650
Db	2035	GAACAGAAGAAAGCTGACGCCCTTGAGGACGCCGACGCTGAGGATCTCACGAATGCAGAT	2094
Qy	651	AlaLeuGluMetSerProLysAspSerGlySerSer-----Leu	663
Db	2095	ATTGCCAAATCAAACTCGTCAACGCCATCATCCGATCTCGACGGAAACCCACAACC	2154
Qy	664	IleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyInAspArgLysLeu	683
Db	2155	ATTGACGGCAGCGGCACCCACAAGTCTGTTCACGGCGGATTTCTTCTAAAGAACCC---	2211
Qy	684	GlyThrLysGluAspLeuAsnGlnAsnValProSerValSerPheTrpArgIleLeuLys	703
Db	2212	-----CCCGAAACAACCTCGGAATACTCATTTATGGACGCTGCTCAAA	2253
Qy	704	-----LeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla	719
Db	2254	TTTGTGTCCTTCAACGCCCTGAATCCGTCATGCTCATCGGCTCTGTCTCTCA	2313
Qy	720	IleIleAsnGlyLysLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle	739
Db	2314	GTGTAGTCTGGTGGCCACCCACGACGACGATGCTATATGCTAAAGCCATCAGCACA	2373
Qy	740	PheThrArgAspGluAspProGluThrLys---ArgGlnAsnSerAsnMetPheSerVal	758
Db	2374	CTCTCGCTCCAGATCACAAATATGCAAGCTTCGACATGATGGGATTTCTGGTCAATTG	2433
Qy	759	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	778
Db	2434	ATGTTCTCTCGTTGGTATCATTCATTCAGTATTATCAGCAGTCAACCAATGGTGGCATTT	2493
Qy	779	GlyLysAlaGlyLuuIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	798
Db	2494	CCGCTATGCTCCGAGACACTTATTCGTCGCGGAGAGCACTGCCTTTCGGACGATACTC	2553
Qy	799	ArgGlnAspValSerThrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg	818
Db	2554	CGTCAAGACATTTGTTCTTTTGAAGGAAGAGATAGCACCGGCGCTCTGACCTCTTC	2613
Qy	819	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	838
Db	2614	CTGTCCACCGAGAGCAAGCATCTCCGGTGTAGCGGTGCTGACTCTAGGCACCATCTTG	2673
Qy	839	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	858
Db	2674	ATGACCTCCAGCACCTTAGGAGCGGTATCATTTATTCCTGGCGATTGGGTGGAAATGG	2733
Qy	859	ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	878
Db	2734	GCCTTAGTTTGTATCTCGGTTGTCGGGTCTCTCGGACGCGGTTCTACCGATTCTAT	2793
Qy	879	MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla	898
Db	2794	ATGTAGCCCAAGTTTCAATCACGCTCCAAAGCTTGCTTATGAGGATCTCGAAACTTTGCT	2853
Qy	899	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu	918
Db	2854	TGCAGGCTACATCGTCTATCCGCACAGTTTGGCTCATTAACCCGGGAAGGAGTGTCTGG	2913
Qy	919	TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle	938
Db	2914	GAGATTTACCATGCCAGCTTTCGACCAAGCAGGACCATGCTCTAATCTCTCTCTGAGG	2973
Qy	939	PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe	958
Db	2974	TCATCCCTGGTATATGGTCTGTCGACGACACTTGTCTTCTTCTGGGTGGCTCGGTTT	3033

Qy	959	ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuVal	978
Db	3034	TGTACCGAGGACACTTTTGGTGCACACGAGTACGATTCATTTCCGCTCTCTTTGTTGT	3093
Qy	979	PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp	998
Db	3094	TTCTCCGAGATCTCTTTGGTGCTCAATCCGCGGCACCGCTTTCTTTTGCACGAC	3153
Qy	999	TyrAlaIysAlaIysValSerAlaAlaHisValIleMetIleGlyLysSerProLeu	1018
Db	3154	ATGGCAAGCGGAAGATCGCGCCGCCGAATCCGACGACTGTTCGACCCGAAAGCCACAA	3213
Qy	1019	IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe	1038
Db	3214	ATTGATAACTGGTCTCAAGAGGCGCAGAGAAGCTCGAAACGGTGGAAAGTGAATCGAATTT	3273
Qy	1039	AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer	1058
Db	3274	AGGAAGCTGCACATTACAGATACCCGACCGCCGACAAAGCCTGCTTCGCCGGCTTGGAC	3333
Qy	1059	LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer	1078
Db	3334	CTGACCGTGAAGCCTGGCAATATGTTGGCTTGTTCGGACCCAGCGTGTGGCAAGAT	3393
Qy	1079	ThrValValGlnLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp	1098
Db	3394	ACCACCATTCGATTGCTTGAGCGCTTTACGATGCGATTGCGGGTCCATCCTTGTGTAT	3453
Qy	1099	GlyLysGluIleLysHisLeuAsnValGlnThrLeuArgAlaHisLeuGlyIleValSer	1118
Db	3454	GGNAGGACATAAGTAACATAATCAACTCCTACCCGACGCTTCTGTCACTGGTCAGC	3513
Qy	1119	GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer	1138
Db	3514	CAGAGCGGACACTTACCAGGCGCACATCAAGGAAACATCTTACTTGGTATTGTCGAA	3573
Qy	1139	ArgValValSerHisGluGluIleMetGlnAlaLysGluAlaAsnIleHisHisPhe	1158
Db	3574	GATGACGTACCGGAAGATCTTGATTAAAGCTTGCAAGGACGCTAATATATACGACTTC	3633
Qy	1159	IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer	1178
Db	3634	ATCATGTCTCCCGAGGCGCTTAATACAGTTGTGGCAGCAGGAGGACGATGTTGCT	3693
Qy	1179	GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu	1198
Db	3694	GCGCGCCAAAAGCAACGTGGCGCATTCGCCGAGCCCTTCTTCGGGATCCCAAAATCCTT	3753
Qy	1199	LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla	1218
Db	3754	CTTCTCGATGAAGCGAGCTCAGCCCTCGACTCCGAGTCAGAAAGGTCGTCACGCGGCT	3813
Qy	1219	LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle	1238
Db	3814	TGTGATCGCGCTGCCCGAGCGCAACCAATCGCGCTTGGCACACCGACTCAGACGATTT	3873
Qy	1239	GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis	1258
Db	3874	CAAAAGCGGACGCTTATCTATGTTTTCGACCAAGGCAAGATCGTCGAAGCGGACGAC	3933
Qy	1259	GlnGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAla	1278
Db	3934	ACCGAACTGGTTCAGAAAAAGGCGCGGTACTACGAGCTGGTCAACTTCGACAGCTTGGC	3993
Qy	1279	Lys 1279	
Db	3994	AAAG 3996	

RESULT 14
US-08-996-545-3
; Sequence 3, Application US/08996545
; Patent No. 5928898

```

: GENERAL INFORMATION:
: APPLICANT: Skatrud, Paul L.
: APPLICANT: de Waard, Maarten A.
: APPLICANT: Peery, Robert B.
: APPLICANT: Andrade, Alan C.
: TITLE OF INVENTION: Multiple Drug Resistance Gene at rD of
: TITLE OF INVENTION: Aspergillus nidulans
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Eli Lilly and Company
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: U.S.
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/996,545
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Webster, Thomas D.
: REGISTRATION NUMBER: 39,872
: REFERENCE/DOCKET NUMBER: X-11766
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-3334
: TELEFAX: 317-276-2763
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4002 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: mRNA
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: US-08-996-545-3

```

Db 1555 ACGACGAAUUAAGAAUUAUAGGCACGCGUCUCAUCGCGCACAAAGUAGCAAGAAUCC 1614
Qy 498 MetAspGlu-----IleGluLysAlaValLysGluAlaAsnAlaTyrAspPhe 513
Db 1615 GAGGAUAAGGUGCCGGAACUCUACGAGAACGCGGCAAAAUAGCGGAUUCUAGCAUUCU 1674
Qy 514 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSer 533
Db 1675 AUUACUGCCUUGCCUGAAGGUUAUGAGACCAUUGUUGGCACGCGUUCUUCUUCUUA 1734
Qy 534 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu 553
Db 1735 GGUGGCAGAAACAGCGCAUUGCAUUGCGCGGCGUGUUAUGAGACCCCAAAAUCCUG 1794
Qy 554 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla 573
Db 1795 CUCCUGGAUGAGCAUCUUGCGCUUGGACACAAAUCCGAAGGCGUGUGUUCACGACGU 1854
Qy 574 LeuAspLysAlaArgLysGlyArgThrIleValIleAlaHisArgLeuSerThrVal 593
Db 1855 UUGGAGAGGCACGUGAAGCGCAACUACUUAUUGAUCGCUACUGCGCUUCCACGAUC 1914
Qy 594 ArgAsnAlaAspValIleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHis 613
Db 1915 AAAACGCGCACACAAUUGUGGUUGUCUUGCAUUGGCAAAAUUGCGAAGAAACUCAC 1974
Qy 614 AspGluLeuMetLysGluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGly 633
Db 1975 GAUGAAUUGUGAACCAGCGGAGCGCUUUCGCAACUUGGAGGCGUUCACGUAUCAU 2034
Qy 634 AsnGluIleGlu-----LeuGluAsnAlaThrGlyGluSerLysSerGluSerAsp 650
Db 2035 GAACAGAAAGAACUGACGCGCUUGGAGGACGCGCGAGGAGUUCACGAAUGCAGAU 2094
Qy 651 AlaLeuGluMetSerProLysAspSerGlySerSer-----Leu 663
Db 2095 AUGGCCAAAAAACAACUGCGUACGCGCAUCUACUACGACGCGGAAACCCACAACC 2154
Qy 664 IleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAspArgLysLeu 683
Db 2155 AUGACGCGACGGCACCCACAGUCUGUUCUCCAGCGCGCAUUCUUAAGAACCC--- 2211
Qy 684 GlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheThrArgIleLeuLys 703
Db 2212 -----CCCGAAACAACUCCGAAAUACUUAUUGAGCAGCGUCGCUCAAA 2253
Qy 704 -----LeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCysAla 719
Db 2254 UUGUUGUUCUUCUUAACCGCCUGAUAUCCCGUACUAGCUACGCGUUGUUCUUCUA 2313
Qy 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
Db 2314 GUGUUAACGUGGUGGCGCAACCCACGACGAGUGCUUAUUGCUAAAGCAUCAGCACA 2373
Qy 740 PheThrArgAspGluAspProGluThrLys---ArgGlnAsnSerAsnMetPheSerVal 758
Db 2374 CUUCGCUCCAGAAACAAUUAAGCAAGCUCUACGUAUGCGGAUUCUGGCUAUG 2433
Qy 759 LeuLeuValLysGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 778
Db 2434 AUGUUCUUGGUGUGUUAUCUUAUGUUAUACGCGACGCAACCAUGGUGGCUAUAU 2493
Qy 779 GlyLysAlaGlyIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 798
Db 2494 GCCGUAGCUCCGAGACAUUAUCGCGCGGAGAGACACUGCCUUCGAGUAUCUC 2553
Qy 799 ArgGlnAspValSerTrpPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 818
Db 2554 CGUACAGACAUUGCUUUGAAGAAAGAAUAGCACCGCGCUCUGACCUUCUUC 2613
Qy 819 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 838
Db 2614 CUGUCCACCGAGACGACUUCUUCGUGUUAUGCGGUGAGCUCUAGCAGCAUCUUG 2673

Qy 839 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 858
Db 2674 AUGACCUCCAGCAGCCUAGGAGCGGUUAUUAUUGCCUGGGAUUGGUGGGAUAUUG 2733
Qy 859 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 878
Db 2734 GCCUUAUGUUAUCUUGCGGUUGUGCGGCAUCGCGCAUCGCGGUUUCACCAUUCUAU 2793
Qy 879 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGlyAlaGlyLysIleAla 898
Db 2794 AUGUAGCCAGUUAUUAUACACGCUCAACGCUUUAUGAGGAGUUCGCAAAUUGCU 2853
Qy 899 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 918
Db 2854 UGCAGGCGUACUUGCUUAUCCGCACAGUUAACCCGGGAAAGGAGUGUGUG 2913
Qy 919 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 938
Db 2914 GAGAUUAACCAUGCCAGCUAGCACACAGGACGAGGACUUAUCUUCUUGUUGAGG 2973
Qy 939 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 958
Db 2974 UCAUCCUGUUAUUGCGUGCGCAGCACUUGUUAUUGCGUGUGCGCGGUGU 3033
Qy 959 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 978
Db 3034 UGGUACGAGGAGCACUUCUUGGUCACACAGUAUGACAUUUUCCGCUUCUUGUUGU 3093
Qy 979 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 998
Db 3094 UUCUCCGAGAUUCUUCUUGGUCUUAUCCGCGGACCGUUCUUCUUGCACCAGAC 3153
Qy 999 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1018
Db 3154 AUGGCAAGGCGAAGAUUGCGCGCCGCGAAUUCGACAGCUUGCGGCGGAAAGCCAA 3213
Qy 1019 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThrPhe 1038
Db 3214 AUUGAAUACUGGUCUUAAGAGGCGGAGAGCUCGAAACGGUGGAGGAGUAAUUGAAUU 3273
Qy 1039 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1058
Db 3274 AGGAACGUGCACUUCAGAUACCCGACCCGCGGAGACGCGUCCGCGCGGCUUGGAC 3333
Qy 1059 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1078
Db 3334 CUGACCGUGAAGCGGACAAUUAUUGCGUUGCGGACCCGCGGUGUGGCAAGAGU 3393
Qy 1079 ThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1098
Db 3394 ACCACAUUGCAUUGCGUAGCGCUUUAACGAGCAUUGCGGCGUCCAUCCUUGUUGAU 3453
Qy 1099 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1118
Db 3454 GGGAGGACAAUAGUAAACUAAUUAUACUCCUACCGCAGCGUUCUGUCACUGGUCAGC 3513
Qy 1119 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1138
Db 3514 CAGGAGCGGACACUGUACCGGCGCACCAUAGGAAACAUUUAUUGGUAUUGUUGAA 3573
Qy 1139 ArgValValSerHisGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1158
Db 3574 GAUGACGUACCGAGAAUUCUUAUUAAGGCUUACGAGGAGGCUUAUUAUUCACGACUUC 3633
Qy 1159 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1178
Db 3634 AUCAUGUCUCGCGGAGGCGUUUAUACAGUUGUGGCGAGGAGGAGGCAUUGUUGU 3693
Qy 1179 GlyGlyGlnLysGluArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1198
Db 3694 GCGGCGCAAAAGCAACGUGGCGCAUUGCCGAGCGCCUUCUUCGCGGAUCCCAAAUCCUU 3753

```
QY 1199 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1218
Db 3754 CUUCUGAAGGAGCGACGCGGACUCCAGUCAGAGAAAGGUGGUCGCGGCGCO 3813
QY 1219 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1238
Db 3814 UUGGAUGCGGUGCGCGGAGCGGACCAACAAUCCGCGUUGCACACCGACGACGAU 3873
QY 1239 GlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1258
Db 3874 CAAGAAGCGGAGGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3933
QY 1259 GlnGlnLeuLeuAlaGlnLysGlyIleThrPheSerMetIleSerValGlnAlaGlyAla 1278
Db 3934 AGCAACUGGUGCCAGAAAGGCGGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 3993
QY 1279 Lys 1279
Db 3994 AAG 3996

RESULT 15
US-09-328-320-1
; Sequence 1, Application US/09328320
; Patent No. 6228615
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene atrd of
; TITLE OF INVENTION: Aspergillus nidulans
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/328,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/996,545
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11766
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-2763
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4002
US-09-328-320-1
Alignment Scores:
```

```
Pred. No.: 1,27e-265 Length: 4002
Score: 2482.00 Matches: 538
Percent Similarity: 59.41% Conservative: 223
Best Local Similarity: 42.00% Mismatches: 458
Query Match: 38.38% Indels: 62
DB: 4 Gaps: 8

US-09-672-725C-4 (1-1280) x US-09-328-320-1 (1-4002)
QY 23 LysLysGluLysLysGluLysLysProThrValSerThrPheAlaMetPheArgTyrSer 42
Db 268 AACACGCGAGTGGAGGAGATCAAGATACTCTCTCTGGTCTCTGGCGGTATGCA 327
QY 43 AsnTrpLeuAspArgLeuThrMetLeuValGlyThrMetAlaAlaIleIleHisGlyAla 62
Db 328 ACAAGATGGATATATCTATCTGTAATCAGTACAACTCTGTCCCATCTCTGCCGCGTG 387
QY 63 AlaLeuProLeuMetMetLeuValPheGlyAsnMetThrAspSerPheAlaAsnAlaGly 82
Db 388 ACTTCCAGAGGATAATGTTATAT-----411
QY 83 IleSerArgAsnLysThrPheProValIleIleAsnGluSerIleThrAsnAsnThrGln 102
Db 412 -----CAA 414
QY 103 HisPheIleAsnHisLeuGluGluMetThrThrTyrAlaTyrTyrTyrSerGlyIle 122
Db 415 ATCTCGTACGACGAGTCTATGATGAATTCAGCAAGACGTACTGTCTGTATACCTC 474
QY 123 GlyAlaGlyValLeuValAlaAlaTyrIleGlnValSerPheTrpCysLeuAlaGly 142
Db 475 GGTATCGGCGAGTGTCTACTCTATGTTAGTACTGTGGCTTCATCTATACCGGAGAA 534
QY 143 ArgGlnIleLeuLysIleArgLysGlnPhePheHisAlaIleMetArgGlnGluIleGly 162
Db 535 CACGCCACGACGAGATCGCGAGTATTACCTTGAGTCTATCTCTCGCCAGAACATTGGC 594
QY 163 TrpPheAspValHisAspValGlyGluLeuAsnThrArgLeuThrAspAspValSerLys 182
Db 595 TATTTTGAATAACTCGGTCGCGGGAAGTACCACCCGTATTAACAGCGCATCAACACCT 654
QY 183 IleAsnGluGlyIleGlyAspLysValGlyMetPhePheGlnSerIleAlaThrPhePhe 202
Db 655 ATCCAGGATGGCATTCGAGAGAAGTCTGCTCACTTTGACTGCGCTCGCGACATCTGTG 714
QY 203 ThrGlyPheIleValGlyPheThrProGlyTrpLysLeuThrLeuValIleLeuAlaIle 222
Db 715 ACAGCATTCATTATCGCCTACGTCAAATACTGGAAGTTGGCTCTAATTTGCAGCTCAACA 774
QY 223 SerProValLeuGlyLeuSerAlaAlaIleTrpAlaLysIleLeuSerSerPheThrAsp 242
Db 775 ATGTGGCCCTCGTCTCACCATGGGCGGTGTTCTCAGTTTATCATCATCAAGTACAGCAA 834
QY 243 LysGluLeuLeuAlaTyrAlaLysAlaGlyAlaValAlaGluGluValLeuAlaAlaIle 262
Db 835 AAGTCGCTTGACAGCTACGGTCAGCGCGCACTGTTGGGAGAGGTCTATCAGTCCATC 894
QY 263 ArgThrValIleAlaPheGlyGlnLysLysGluLeuGluArgTyrAsnLysAsnLeu 282
Db 895 AGAAATGCCACAGCGCTTTGGCACCACCAAGCAAGCTTGGCAAGCAGTATGAGTGCACCTA 954
QY 283 GluGluAlaLysArgIleGlyIleLysLysAlaIleThrAlaAsnIleSerIleGlyAla 302
Db 955 GACGAAGCTGAGAAATGGGGAACAAAGAACCCAGATTGTGATGGGTTCATGATGGGCGC 1014
QY 303 AlaPheLeuIleTyrAlaSerTyrAlaLeuAlaPheTrpTyrGlyThrSerLeuVal 322
Db 1015 ATGTTTGGCTTATGCTACTCGAAGTACGGTCTTGGCTCTGATGGGTCTCGTTCTCTG 1074
QY 323 LeuSerSerGluTyrThrIleGlyGlnValLeuThrValPhePheSerValLeuIleGly 342
Db 1075 GTAGATGTCGAGTCGATGGGTGATATTCACAGTTCCTCAGATTCCTGATGGCATCTTGCAT 1134
```

QY 343 AlaPheSerIleGlyGlnAlaSerProSerIleGluAlaPheAlaAsnAlaArgGlyAla 362
Db 1135 TCGTCTCTTGGGAAACGTTAGTCCAAATGCTCAAGCATTTACAAACGCTGTGGCCGG 1194
QY 363 AlaTyrGluIlePheLysIleLeuAsnLysProSerIleAspSerTyrSerLysSer 382
Db 1195 GCCGCAAGATATTTGGAACGATCGATCGCCAGTCCCATATAGATCCATATTCGAACGAA 1254
QY 383 GlyHisLysProAspAsnIleLysGlyAsnLeuGluPheLysAsnValHisPheSerTyr 402
Db 1255 GGAAGACGCTCGACATTTTGGGCCACATTTAGGCGCCATTTAGCAATGCTCAAGCATATTTAC 1314
QY 403 ProSerArgLysGluValLysIleLeuLysGlyLeuAsnLysValGlnSerGlyGln 422
Db 1315 CCATCTAGACCGAGGTACCGTCACTGAGGATGTTCTCTGTCAATGCCCGCTGAAAA 1374
QY 423 ThrValAlaLeuValGlyAsnSerGlyCysGlyLysSerThrThrValGlnLeuMetGln 442
Db 1375 ACAACGCTTTAGTCGGCCCTCTGGCTCTGGAAAAAGTACGGTGGCTTGGTTGAG 1434
QY 443 ArgLeuTyrAspProThrAspGlyMetValCysIleAspGlyGlnAspIleArgThrIle 462
Db 1435 CGATTCTACATGCTGTTCGGGTACGGTTCGTGGATGGCCATGACATCAAGACCTC 1494
QY 463 AsnValArgHisLeuArgGluIleThrGlyValValSerGlnGluProValLeuPheAla 482
Db 1495 AATCTCGCTGGCTTCGCCAACAGATCTCTTGGTTAGCCAGGAGCTGTTCTTTTGGC 1554
QY 483 ThrThrIleAlaGluAsnIleArgTyrGly- 497
Db 1555 ACGACGATTTATAAGAATATTAGCAGCGTCTCATCGGCACAAAGTACGAGAAATGAATCC 1614
QY 498 MetAspGlu- 513
Db 1615 GAGATTAAGTCGCGGAACTCATCGAAGCCGCGCAAAATGCGGAATGCTCATGACTTT 1674
QY 514 IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSer 533
Db 1675 ATTACTGCTTGCCTGAAGGTTATGAGCAATGTTGGCAGCGTGGCTTCTCTCTTCA 1734
QY 534 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu 553
Db 1735 GGTGGCCAGAAACAGCGCATTCGAATCGCCGCGCTGTTTGTAGTACCCCAAAATCCCTG 1794
QY 554 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValGlnValAla 573
Db 1795 CTCCTGATGAAGCTACTTCGGCTTGACACAAATCCGAAGCGTGGTTCGAAGCAGCT 1854
QY 574 LeuAspLysAlaArgLysGlyArgThrThrIleValIleAlaHisArgLeuSerThrVal 593
Db 1855 TTGGAGAGGCGAGCTGAAGCGCGAATCTATTGTGATCGCTCATCGCCTTTCACGATC 1914
QY 594 ArgAsnAlaAspValIleAlaGlyPheAspGlyValIleValGluLysGlyAsnHis 613
Db 1915 AAAACGGCGCACACATTTGGTTCGGTCAATGGCAAAATTCCTGAACGAAGCACTCAC 1974
QY 614 AspGluLeuMetLysGluLysIleTyrPheLysLeuValThrMetGlnThrArgGly 633
Db 1975 GATCAATTTGTTGACCGCGAGCGCTTATCGCAAACTTGTGGAGGCTCAACGTATCAAT 2034
QY 634 AsnGluIleGlu- 650
Db 2035 GAACGAGAAGGAGTACGCGCTTGGAGGACCGCGAGCTGAGGATCTCAGAAATGAGAT 2094
QY 651 AlaLeuGluMetSerProLysAspSerGlySerSer- 663
Db 2095 ATTGCCAAATCAAACTGCTCAAGCGCATCATCGATCTCGACGGAACCCACAACC 2154
QY 664 IleLysArgArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAspArgLysLeu 683
Db 2155 ATTGACGCGACGCGCACCCCAAGCTGTTTCCAGCGCGATCTTCTCTAAAGACCC- 2211
QY 684 GlyThrLysGluAspLeuAsnGluAsnValProSerValSerPheThrArgIleLeuLys 703

Db 2212 -----CCGAAACAACCTCCGAAATFACATATATGAGCGCTGCTCAA 2253
QY 704 -----LeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCysAla 719
Db 2254 TTTGTTGCTTCTTCAACCGCCCTGAAATCCGTACATGCTCATCGTCTTGTCTTCTCA 2313
QY 720 IleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIle 739
Db 2314 GTGTTAGTGTGTGGCCAAACCCACGACGAGTGTATATGCTAAAGCATCAGCACA 2373
QY 740 PheThrArgAspGluAspProGluThrLys-ArgGlnAsnSerAsnMetPheSerVal 758
Db 2374 CTCGCTCCAGAAATCACAAATATAGCAAGCTTCGACATGATCGGATTTCTGTCATTG 2433
QY 759 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlnPheThrPhe 778
Db 2434 ATGTTCTTCGTGGTGTGATCATTCAGTTATACAGCAGTCAACCAATGTTGCTGCAAT 2493
QY 779 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 798
Db 2494 GCGGTATGCTCCGAGAGACTTATTCTCGCGGAGAGCACTGCTTTCGAGCATACTC 2553
QY 799 ArgGlnAspValSerTyrPheAspAspProLysAsnThrThrGlyAlaLeuThrThrArg 818
Db 2554 CGTCAAGACATTTGCTTTTGACAAGAAAGAAATAGCACCGCGCTCTGACCTCTTTC 2613
QY 819 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 838
Db 2614 CTGTCACCGGAGACGACATCTCCGGTGTAGCGGTGTGACTCTAGGCAGCATCTG 2673
QY 839 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTyrGlnLeu 858
Db 2674 ATGACCTCCAGACGCTAGGAGCGCTATCATTTATGCGCTGGCGATTTGGGTGAAAT 2733
QY 859 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 878
Db 2734 GCCTTAGTTGTATCTCGGTGTGCGGTCTCTCGGTGATCGGTCTTCTACCGATCTAT 2793
QY 879 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 898
Db 2794 ATGCTAGCCCACTTTCAATCAGCTCCAGCTTGTGTTATGAGGATCTGCAAACTTTGCT 2853
QY 899 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrArgGluGlnLysPheGlu 918
Db 2854 TCGGAGGCTACATCTCTATCCGCACAGTTGCGTCAATTAACCGGGAAGGATGCTGG 2913
QY 919 TyrMetTyrAlaGlnSerLeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIle 938
Db 2914 GAGATTTACCATGCCAGCTTGACGCAAGCAGGACGACGCTTAATCTCTGCTTGAGG 2973
QY 939 PheGlyValSerPheSerIleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPhe 958
Db 2974 TCATCCCTGTATATGCGTTCGTCGAGGCACTTGTTTTCTTCGCTTTCGCTCGGGTTT 3033
QY 959 ArgPheGlyAlaTyrLeuValAlaAsnGluPheMetAsnPheGlnAspValLeuVal 978
Db 3034 TGGTACGGAGGACACTTCTTGGTCCACGAGTATGACATTTCCGCTTCTTGTGTGT 3093
QY 979 PheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPheAlaProAsp 998
Db 3094 TTCTCCGAGATTTCTTTGGTCTCAATCCGGGCGACGCTGTTTCTTCTTTCGTCACGAC 3153
QY 999 TyrAlaLysAlaLysValSerAlaAlaHisValIleMetIleIleGluLysSerProLeu 1018
Db 3154 ATGGCGAAGCGGAAGATGCGGCCCGGAAATTCGACGACTGTTCCAGCCGAAGCCACA 3213
QY 1019 IleAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsnValThr 1038
Db 3214 ATGATACTGCTGAAGAGGCGGAGAGCTCGAAGCGTGAAGGTGAATCGAATTT 3273
QY 1039 AsnGluValValPheAsnTyrProThrArgProAspIleProValLeuGlnGlyLeuSer 1058

Db 3274 AGGAACGTGCACTTCAGATACCGACCGCCAGAACAGCTGTCTCGCGGGCTTGGAC 3333
QY 1059 LeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCysGlyLysSer 1078
Db 3334 CTGACCGTGAAGCCTGGACAATATGTTGCGGTGTGCGACCCACGGTGTGGCAAGAGT 3393
QY 1079 ThrValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerValLeuIleAsp 1098
Db 3394 ACCACCATTCATTCGTGAGCGCTTTACGATGCGATTGCGGGTCCATCTCTGTTGAT 3453
QY 1099 GlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGlyIleValSer 1118
Db 3454 GGAAGGACATAAGTAATAATCACTACCTACCGAGCTTCTGTCTACTGTCAGC 3513
QY 1119 GlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSer 1138
Db 3514 CAGGACCGACACTGTACCGGGCCACCAAGGAACATCTTACTTGGTATTGTGCGAA 3573
QY 1139 ArgValValSerHisGluGluIleMetGlnAlaAlaLysGluAlaAsnIleHisHisPhe 1158
Db 3574 GATGACGTACCGGAAGAATCTTGATTAAAGCTTGCAAGGACGCTAATATCTACGACTTC 3633
QY 1159 IleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThrGlnLeuSer 1178
Db 3634 ATCATGTGCTCCCGAGGGCTTTAATACAGTTGTGCGACAGGGAGGCATGTTGCT 3693
QY 1179 GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnProHisIleLeu 1198
Db 3694 GCGGCCAAGCAACAGTGTGGCCATTGCGGAGCCCTTCTCGGGATCCCAAAATCCTT 3753
QY 1199 LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValValGlnGluAla 1218
Db 3754 CTTCGTGATGAAGCGAGCTCAGCCCTCGACTCCGAGTCAGAAAAGTCTCCAGGGCGCT 3813
QY 1219 LeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeuSerThrIle 1238
Db 3814 TTGGATCCCGCTGCCCGAGGCGCAACCACTGCGCGTTCACACCCAGCTCAGCAGGATT 3873
QY 1239 GlnAsnAlaAspLeuIleValPheGlnAsnGlyLysValLysGluHisGlyThrHis 1258
Db 3874 CAAAGCGGACGTTATCTATGTTTCGACCAAGGCAAGATCGTCGAAAGCGGAACGCAC 3933
QY 1259 GlnGlnLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValGlnAlaGlyAla 1278
Db 3934 AGCGAACTGGTCCAGAAAAAGGCGCGTACTACGAGCTGGTCAACTGTCAGAGCTTGGGC 3993
QY 1279 Lys 1279
Db 3994 AAG 3996

Search completed: November 6, 2002, 19:16:20
Job time : 383.51 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:33:09 ; Search time 20.0473 Seconds
(without alignments)
6135.192 Million cell updates/sec

Title: US-09-672-725C-4
Perfect score: 6467
Sequence: 1 MDPEGGRKGAENFWMGK.....LLAQGIYFSMISVQAGAKR 1280
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5859.5	90.6	1280	1 DVHU1	multidrug resist
2	5650	87.4	1276	1 DVHY1C	multidrug resist
3	5631.5	87.1	1276	2 A34786	multidrug resist
4	5304.5	82.0	1276	1 DVMS1	multidrug resist
5	5236	81.0	1277	2 JH0502	multidrug resist
6	4982	77.0	1104	1 DVMS1A	p-glycoprotein - r
7	4898	75.7	1279	1 DVHU3	multidrug resist
8	4850.5	75.0	1276	1 DVMS2	multidrug resist
9	4810	74.4	1281	2 I48123	multidrug resist
10	4796.5	74.2	1278	2 S41646	p-glycoprotein iso
11	4422.5	68.4	1287	2 S55692	p-glycoprotein - r
12	3201	49.5	1321	2 T42228	multidrug resist
13	3197	49.4	1321	2 T42842	p-glycoprotein sis
14	2867	44.3	1294	2 T19982	multidrug resist
15	2810	43.5	1275	2 T31073	hypothetical prote
16	2809	43.4	1289	2 D87789	multidrug resist
17	2740	42.4	1321	2 T23476	protein C34G6.4 [i
18	2738	42.3	1321	2 S27337	hypothetical prote
19	2675.5	41.4	655	1 DVHY2C	multidrug resist
20	2635.5	40.8	1283	2 A47377	multidrug resist
21	2610	40.4	1292	2 T47807	p-glycoprotein hom
22	2594	40.1	1278	2 E86155	probable ABC trans
23	2588.5	40.0	1286	2 T02187	probable ABC trans
24	2561.5	39.6	1302	2 A41249	multidrug resist
25	2533	39.2	1229	2 D85023	p-glycoprotein-lik
26	2530	39.1	1229	2 T52319	p-glycoprotein-lik
27	2519	39.0	1230	2 E85023	probable p-glycopr
28	2493	38.5	1302	2 B41249	multidrug resist
29	2479	38.3	1229	2 F86155	probable ABC trans

30	2434.5	37.6	1323	2 H85202	hypothetical prote
31	2432.5	37.6	1408	2 T43261	multidrug resist
32	2431	37.6	1310	2 S30328	multidrug resist
33	2423	37.5	1286	2 A42150	p-glycoprotein ppg
34	2408.5	37.2	1266	2 T22090	hypothetical prote
35	2400.5	37.1	1268	2 T22094	hypothetical prote
36	2386	36.9	1302	2 S30327	multidrug resist
37	2347.5	36.3	1254	2 S27338	p-glycoprotein C -
38	2327.5	36.0	1222	2 T14805	hypothetical prote
39	2301	35.6	1233	2 T04251	p-glycoprotein 2 -
40	2299.5	35.6	1245	2 G86404	probable p-glycopr
41	2287.5	35.4	1254	2 T30855	multidrug resist
42	2267.5	35.1	1362	2 T41534	leptomycin B resis
43	2267	35.1	1307	2 T30882	multidrug resist
44	2265.5	35.0	1318	2 T21266	hypothetical prote
45	2250.5	34.8	1247	2 F86405	probable p-glycopr

ALIGNMENTS

RESULT 1

DVHU1

multidrug resistance protein 1 - human

N:Alternate names: P-glycoprotein 1

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1990 #sequence_revision 18-Aug-1995 #text_change 19-Jan-2001

C:Accession: A34914; PS0162; S15500; S43838; I52238; I65204

R:Chen, C.; Clark, D.; Ueda, K.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.

J. Biol. Chem. 265, 506-514, 1990

A:Title: Genomic organization of the human multidrug resistance (MDR1) gene and origin

A:Reference number: A34914; MUID:90094448

A:Accession: A34914

A:Molecule type: DNA

A:Residues: 1-1280 <CHE>

A:Cross-references: GB:M29447; GB:J05168; NID:g187496; PIDN:AAA59576.1; PID:g386862

R:Kioaka, N.; Yamano, Y.; Komano, T.; Ueda, K.

submitted to JIPID, April 1991

A:Reference number: PS0162

A:Accession: PS0162

A:Molecule type: DNA

A:Residues: 1-22 <KIO>

R:Kioaka, N.; Yamano, Y.; Komano, T.; Ueda, K.

submitted to the EMBL Data Library, April 1991

A:Description: Transcriptional regulation of multidrug resistance gene (MDR1) expressed

A:Reference number: S15500

A:Accession: S15500

A:Molecule type: DNA

A:Residues: 1-22, 'R' <K12>

A:Cross-references: EMBL:X58723; NID:g34522; PIDN:CAA41558.1; PID:g34523

R:Chen, C.; Chin, J.E.; Ueda, K.; Clark, D.P.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.

A:Cross-references: GB:M37724; NID:g183537; PIDN:AAA88047.1; PID:g553314
A:Accession: I55204
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 800-856 <RE2>
A:Cross-references: GB:M37725; NID:g183538; PIDN:AAA88048.1; PID:g553315
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: GDB:PGY1; MDR1
A:Cross-references: GDB:120712; OMIM:171050
A:Map position: 7q21-7q21
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; phosphoprotein;
F:1-638,635-1280/Region: duplication
F:49-350/Domain: hydrophobic <HB1>
F:351-637/Domain: hydrophilic <HL1>
F:410-604/Domain: ATP-binding cassette homology <ABC1>
F:427-434/Region: nucleotide-binding motif A (P-loop)
F:551-555/Region: nucleotide-binding motif B
F:638-708/Domain: linker <LIN>
F:709-993/Domain: hydrophobic <HB2>
F:994-1280/Domain: hydrophilic <HL2>
F:1053-1249/Domain: ATP-binding cassette homology <ABC2>
F:1070-1077/Region: nucleotide-binding motif A (P-loop)
F:1196-1200/Region: nucleotide-binding motif B
F:91.94.99/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:433/Binding site: ATP (Lys) #status predicted
F:661,667,671/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status exp
F:667,671,683/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status
F:1076/Binding site: ATP (Lys) #status predicted
Query Match 90.6%; Score 5859.5; DB 1; Length 1280;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 1159; Conservative 59; Mismatches 60; Indels 3; Gaps 3;
Qy 1 MDPEGGKSGA-EKNFWMGKSKKKEKKKPTVSTFAMFRYSNWLDRMLVGLTMAAII 59
Db 1 MDLEGRNGAKKKNPKLNKSEKKKPTVSTFAMFRYSNWLDRMLVGLTMAAII 60
Qy 60 HGAALPLMLVFGNWTDSFANAGISRNKTPPVIIINESITNNTQHFNLHLEEMTTVAYY 119
Db 61 HGAGLPLMLVFGNWTDSFANAG-NLEDLMSNTNRSIDINDTGFNM-LEEDMTRVAYY 118
Qy 120 SGIGAGVLVAAYIQVSWFCLAAQRQILKIRKQFFHAIMRQEIWFVDVHVGELNRLTDD 179
Db 119 SGIGAGVLVAAYIQVSWFCLAAQRQILKIRKQFFHAIMRQEIWFVDVHVGELNRLTDD 178
Qy 180 VSKINEGIDKGMFFOSIAFTFTGFTVPGWKLTLVLAISPVGLSAAWAKILSS 239
Db 179 VSKINEGIDKGMFFOSIAFTFTGFTVPGWKLTLVLAISPVGLSAAWAKILSS 238
Qy 240 FTDKELLAYAKAGAAVEVLAARTVTAFGQKKELERYNKNLEAKRIGIKKAITANIS 299
Db 239 FTDKELLAYAKAGAAVEVLAARTVTAFGQKKELERYNKNLEAKRIGIKKAITANIS 298
Qy 300 IGAFLTIYALAFWGTSLVLSSEYITGQVLTVFVSSVIGAFSTGQASPSIEAFANA 359
Db 299 IGAFLTIYALAFWGTSLVLSSEYITGQVLTVFVSSVIGAFSTGQASPSIEAFANA 358
Qy 360 RGAAYEIKIINDKPSIDSKSHKPDNIKGNLFKNVHFSYPSRKEVKILGLNLKVQ 419
Db 359 RGAAYEIKIINDKPSIDSKSHKPDNIKGNLFKNVHFSYPSRKEVKILGLNLKVQ 418
Qy 420 SGQTVLVGNSGCKSTTVQLMQLRXYDPTDGMVCDIGQDINTINVRHLREITGVVSEPV 479
Db 419 SGQTVLVGNSGCKSTTVQLMQLRXYDPTDGMVCDIGQDINTINVRHLREITGVVSEPV 478
Qy 480 LFATTIAENIRYGRNVTMDIEKAVKANNAYDFIMKLPKFDTLVGERGARLSGGQQR 539
Db 479 LFATTIAENIRYGRNVTMDIEKAVKANNAYDFIMKLPKFDTLVGERGARLSGGQQR 538
Qy 540 TAIARALVRNPKILLDLDEATSDALDTESEAVVQVQALDKARKGRITVIAHRLSTVRNADVI 599

Db 539 TAIARALVRNPKILLDLDEATSDALDTESEAVVQVQALDKARKGRITVIAHRLSTVRNADVI 598
Qy 600 AGFDGVIYKGNHDELMKEGIYFKLVMTQTRKNSNMFSLVLGLGIISFTFFLQGFTEG 659
Db 599 AGFDGVIYKGNHDELMKEGIYFKLVMTQTRKNSNMFSLVLGLGIISFTFFLQGFTEG 658
Qy 660 GSSLIKRSTRSRSHAPQGDQRKLTGKEDLNENVPVSFWRILKLNSTWPFYVVGIFCA 719
Db 659 RSSLIRKSTRSRVSGSQADRKLTKEALDESIPPVSWRIMKLNLTWPFYVVGIFCA 718
Qy 720 IINGLOPAPSIIFSRIGIIFTRDDETPETKRONSMFSLVLGLGIISFTFFLQGFTEG 779
Db 719 IINGLOPAPSIIFSRIGIIFTRDDETPETKRONSMFSLVLGLGIISFTFFLQGFTEG 778
Qy 780 KAGELTKRLRYMFRSMRLQDVSWDFDDPKNTTGALTTRLANDAAQVKGASRLAVITQ 839
Db 779 KAGELTKRLRYMFRSMRLQDVSWDFDDPKNTTGALTTRLANDAAQVKGASRLAVITQ 838
Qy 840 NIANLGTGIIISLIYQWLTLLLAIVPIAIAAGVVEKMLSGOALKDKELEGAGKIAT 899
Db 839 NIANLGTGIIISLIYQWLTLLLAIVPIAIAAGVVEKMLSGOALKDKELEGAGKIAT 898
Qy 900 EAIENFRVVSITREKFEYMYAOSLOVPYRNSLRKAHIFGVSFSTQAMMYFSYAGCFR 959
Db 899 EAIENFRVVSITREKFEYMYAOSLOVPYRNSLRKAHIFGVSFSTQAMMYFSYAGCFR 958
Qy 960 FGAYLVANEFMNFQDLVLFSAIVFGAMAVQVSSFPADYAKAKVSAHVIMITEKSPLI 1019
Db 959 FGAYLVANEFMNFQDLVLFSAIVFGAMAVQVSSFPADYAKAKVSAHVIMITEKSPLI 1018
Qy 1020 DSYSPHGLKPNTEGVNFVFNPRDIPVQLGSLVLEKVGOTLALVSGSCGCKST 1079
Db 1019 DSYSPHGLKPNTEGVNFVFNPRDIPVQLGSLVLEKVGOTLALVSGSCGCKST 1078
Qy 1080 VVQLLERYDPLAGSLVDGKEIKHLNVQWLRHAGLIGVSOBPIFLFDCSIAENIAYGNSR 1139
Db 1079 VVQLLERYDPLAGSLVDGKEIKHLNVQWLRHAGLIGVSOBPIFLFDCSIAENIAYGNSR 1138
Qy 1140 VYSHEEIMQAAKEANHHFIETLPEKYNTVRGDKGTQLSGGQKQRIATARALVRPHILL 1199
Db 1139 VYSHEEIMQAAKEANHHFIETLPEKYNTVRGDKGTQLSGGQKQRIATARALVRPHILL 1198
Qy 1200 IDEATSDALDTESEKVVQVQALDKARKGRITVIAHRLSTIQNALDVLVFNQGVKVEHGHQ 1259
Db 1199 IDEATSDALDTESEKVVQVQALDKARKGRITVIAHRLSTIQNALDVLVFNQGVKVEHGHQ 1258
Qy 1260 QLLAQKGIYFSMISVQAGAKR 1280
Db 1259 QLLAQKGIYFSMISVQAGAKR 1279
RESULT 2
DVHYLC
Multidrug resistance protein 1 - Chinese hamster
N:Alternate names: p-glycoprotein pgp1
C:Species: Crictetus griseus (Chinese hamster)
C:Date: 31-Dec-1990 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A38696; B38696; A27126; S33768; I52823
R:Devine, S.E.; Hussain, A.; Davide, J.P.; Melera, P.W.
J. Biol. Chem. 266, 4545-4555, 1991
A:Title: Full length and alternatively spliced pgp-1 transcripts in multidrug-resista
A:Reference number: A38696; MUID:91154265
A:Accession: A38696
A:Molecule type: mRNA
A:Residues: 1-1276 <DE>
A:Cross-references: GB:M59253; NID:g191154; PIDN:AAA37004.1; PID:g191155
A:Accession: C38696
A:Molecule type: mRNA
A:Residues: 108-1276 <DE1>
A:Cross-references: GB:M59254; NID:g191156; PIDN:AAA37005.1; PID:g191157
A:Experimental source: clone ADX185
A:Accession: B38696

A:Molecule type: mRNA
A:Residues: 1-32, 771-1276 <DE2>
A:Cross-references: GB:M59252; NID:g191152; PIDN:AAA37003.1; PID:g191153
A:Experimental source: Clone ADX124
R:Endicott, J.A.; Juranka, P.F.; Sarangi, F.; Gerlach, J.H.; Deuchars, K.L.; Ling, V.
Mol. Cell. Biol. 7, 4075-4081, 1987
A:Title: Simultaneous expression of two P-glycoprotein genes in drug-sensitive Chinese hamster ovary cells
A:Reference number: A27126; MUID:88122132
A:Accession: A27126
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 706-1276 <END>
A:Cross-references: GB:M17897; NID:g191158; PIDN:AAA37006.1; PID:g191159
R:Zastawny, R.L.; Ling, V.
Biochim. Biophys. Acta 1173, 303-313, 1993
A:Title: Structural and functional analysis of 5' flanking and intron 1 sequences of the human multidrug resistance gene
A:Reference number: S33768; MUID:93305724
A:Accession: S33768
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-21 <ZAS>
A:Cross-references: EMBL:L03286
R:Teeter, L.D.; Eckersberg, T.; Tsai, Y.; Kuo, M.T.
Cell Growth Differ. 2, 429-437, 1991
A:Title: Analysis of the Chinese hamster P-glycoprotein/multidrug resistance gene pgp1 and its promoter
A:Reference number: I52823; MUID:92088970
A:Accession: I52823
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-21 <RES>
A:Cross-references: GB:S81975; NID:g240862
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell lines and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: pgp1
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: alternative splicing; ATP; duplication; glycoprotein; nucleotide binding; P-glycoprotein; multidrug resistance protein
F:407-601/Domain: ATP-binding cassette homology <ABC1>
F:424-431/Region: nucleotide-binding motif A (P-loop)
F:548-552/Region: nucleotide-binding motif B
F:1050-1246/Domain: ATP-binding cassette homology <ABC2>
F:1067-1074/Region: nucleotide-binding motif A (P-loop)
F:1193-1197/Region: nucleotide-binding motif B
F:87,91,96/Binding site: carboxylate (Asn) (covalent) #status predicted
F:430/Binding site: ATP (Lys) #status predicted
F:1073/Binding site: ATP (Lys) #status predicted
Query Match 87.4%; Score 5650; DB 1; Length 1276;
Best Local Similarity 86.6%; Pred. No. 0;
Matches 1109; Conservative 81; Mismatches 85; Indels 6; Gaps 2;
QY 1 MDPGGKGSABKNFWMGKKKKKKKPTVSTFAMFRYNNWLDRLVMTAAIIH 60
Db 1 MEFEEDSGKDKNFLLKMGKKKKKKKPVVSTFMRVAGWLDRLVMTAAIIH 60
QY 61 GAALPLMLVFGNMTDSFANAG-ISRNTFTPVIIINESITNNTQHFNLHEEMTYAAYY 119
Db 61 GVALPLMLVFGNMTDSFASVGNIPNAT-----NNATQVNASDIFGKLEEMTYAAYY 115
QY 120 SGIGAGLVAAVIOVSFWCLAGROILKIRKOFFHAIHQEIGWDFDVGELNRLTDD 179
Db 116 TGIGAGLVAVIOVSFWCLAGROILKIRKOFFHAIHQEIGWDFDVGELNRLTDD 175
QY 180 VSKINEGIDKVMFPQSITATFTFTGTPGKLTILVILAISSVPLGISAIWAKILSS 239
Db 176 VSKINEGIDKVMFPQATFTFTGTPGKLTILVILAISSVPLGISAIWAKILSS 235
QY 240 FTDKELLAYAKAGAAEVLAIIRTVIAPGGOKKELRYNNKLEAKRIGIKKAITANIS 299
Db 236 FTDKELQAYAKAGAAEVLAIIRTVIAPGGOKKELRYNNKLEAKRIGIKKAITANIS 295
QY 300 IGAAFLLIYASALAFWGTSTLVSSEYTIQGLTVFVSFLIGAFSIGQASPNIEAFANA 359
Db 300 IGAAFLLIYASALAFWGTSTLVSSEYTIQGLTVFVSFLIGAFSIGQASPNIEAFANA 355

Db 296 MGAFLLIYASALAFWGTSTLVSKEYSIGQVLTFFFAVLIAPFSGIQASPNIEAFANA 355
QY 360 RGAAYEIFKIIDNKPSIDSYSGKHKPDNTKGNLEEFKNHVSFSPSKVEKILKGLNKLKQV 419
Db 356 RGAAYEIFNIIDNKPSIDSFSGKNGYFDPNKGLEFKNHVSFSPSKVDQVILKGLNKLKQV 415
QY 420 SGQTVLVNMGSGCGKSTTVOLMQLYDPTDGMVCDIGQDRTINVRHLREITGVVSEQPV 479
Db 416 SGQTVLVNMGSGCGKSTTVOLMQLYDPTDGMVCDIGQDRTINVRHLREITGVVSEQPV 475
QY 480 LPATTIAENIRYGRNVTMDTEKAVKANAYDFTMKLPNKEDTLVGERGARGSGQKOR 539
Db 476 LPATTIAENIRYGRNVTMDTEKAVKANAYDFTMKLPNKEDTLVGERGARGSGQKOR 535
QY 540 IATARALVRNPKILLDEATSALDTESEAVVQVVALDKARKGRTTIVIAHRLSTVRNADVI 599
Db 536 IATARALVRNPKILLDEATSALDTESEAVVQVVALDKARKGRTTIVIAHRLSTVRNADVI 595
QY 600 AGFDDGVIVKGNHDELMKEGIYFKLVMTQRTGNEIELEENATGESKSESDALEMSPKDS 659
Db 596 AGFDDGVIVKGNHDELMKEGIYFKLVMTQRTGNEIELEENATGESKSESDALEMSPKDS 655
QY 660 GSSLIKRRSTRRSIIHAPQODRKLTGKEDLNENVPVSFWRLKLNSTEWPFVYVGFICA 719
Db 656 GSSLIKRRSTRRSIIHAPQODRKLTGKEDLNENVPVSFWRLKLNSTEWPFVYVGFICA 715
QY 720 IINGLQPAFSIIFRSIIIGITRDEDPETKRONSMFSLVFLVGLIISFIITFFLQGFIFG 779
Db 716 IINGLQPAFSIIFRSIIIGITRDEDPETKRONSMFSLVFLVGLIISFIITFFLQGFIFG 775
QY 780 KAGEILTKRLYVWFRMLRQDVSWDFDPKNTGTALTRLANDAAQVKGATGSLRAVITQ 839
Db 776 KAGEILTKRLYVWFRMLRQDVSWDFDPKNTGTALTRLANDAAQVKGATGSLRAVITQ 835
QY 840 NIANLGTGIIISLIYGVWLTLLLLAIPIAIVPIAIVAGVEMKLSGQALKDKKELEGAKTAT 899
Db 836 NIANLGTGIIISLIYGVWLTLLLLAIPIAIVPIAIVAGVEMKLSGQALKDKKELEGAKTAT 895
QY 900 EAIENFRVVSILTRQKFEYMYAQSLOVYPYRNSLKAHIFGVSTSIQAMMYFYAGCFR 959
Db 896 EAIENFRVVSILTRQKFEYMYAQSLOVYPYRNSLKAHIFGVSTSIQAMMYFYAGCFR 955
QY 960 FGAYLVANEFNFDVLLVSAIVFGAMAVGVSSFPADYAKAKVSAHVIMIEKSPLI 1019
Db 956 FGAYLVANEFNFDVLLVSAIVFGAMAVGVSSFPADYAKAKVSAHVIMIEKSPLI 1015
QY 1020 DSYSPHGLKPNLEGNVTNFNEVFNYPTRDPVPLQGLSLEVKKGTALVYSSGCGKST 1079
Db 1016 DSYSPHGLKPNLEGNVTNFNEVFNYPTRDPVPLQGLSLEVKKGTALVYSSGCGKST 1075
QY 1080 VVOLLERYDPLAGSLVDGKEIKHLNVQWLAHLAGVVSQEPILFDCSIAENIAYGNSR 1139
Db 1076 VVOLLERYDPLAGSLVDGKEIKHLNVQWLAHLAGVVSQEPILFDCSIAENIAYGNSR 1135
QY 1140 VVSHEEIMQAARAKNIHFIETLPEKYNTRVGDGKGTGSGGOKORIAIARALVQPHILL 1199
Db 1136 VVSHEEIMQAARAKNIHFIETLPEKYNTRVGDGKGTGSGGOKORIAIARALVQPHILL 1195
QY 1200 LDEATSDALTESEKVKVQVQALDKAREGRTCIIVIAHRLSTIQNADLIIVVFQNGKVEHGTQ 1259
Db 1196 LDEATSDALTESEKVKVQVQALDKAREGRTCIIVIAHRLSTIQNADLIIVVFQNGKVEHGTQ 1255
QY 1260 QLLAQKGIYFSMISVQAGAKR 1280
Db 1256 QLLAQKGIYFSMISVQAGAKR 1276

RESULT 3

A34786

multidrug resistance protein 1a - mouse

C:Species: Mus musculus (house mouse)

C:Date: 13-Jul-1990 #sequence_revision 13-Jul-1990 #text_change 02-Feb-2001

C:Accession: A34786; A35671

R:Devault, A.; Gros, P.
Mol. Cell. Biol. 10, 1652-1663, 1990
A:Title: Two members of the mouse mdr gene family confer multidrug resistance with over
A:Reference number: A34786; MUID:90205845
A:Accession: A34786
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1276 <DEV>
A:Cross-references: GB:M30697; NID:g199111; PIDN:AAA39517.1; PID:g387429
R:Hsu, S.I.H.; Cohen, D.; Kirschner, L.S.; Lothstein, L.; Hartstein, M.; Horwitz, S.B.
Mol. Cell. Biol. 10, 3596-3606, 1990
A:Title: Structural analysis of the mouse mdr1a (P-glycoprotein) promoter reveals the ba
A:Reference number: A35671; MUID:90287150
A:Accession: A35671
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-525, 'HV', 528-938, 'S', 940-1035, 'V', 1037-1276 <HSU>
A:Cross-references: GB:M33581; NID:g199104; PIDN:AAA39514.1; PID:g387427
A:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; membrane protein; nucleotide binding; P-loop
F:406-600/Domain: ATP-binding cassette homology <ABC1>
F:423-431/Region: nucleotide-binding motif A (P-loop)
F:547-551/Region: nucleotide-binding motif B
F:1049-1245/Domain: ATP-binding cassette homology <ABC2>
F:1066-1074/Region: nucleotide-binding motif A (P-loop)
F:1192-1196/Region: nucleotide-binding motif B
F:1229/Binding site: ATP (Lys) #status predicted
F:1072/Binding site: ATP (Lys) #status predicted

Query Match 87.1%; Score 5631.5; DB 2; Length 1276;
Best Local Similarity 86.7%; Pred. No. 0;
Matches 1111; Conservative 78; Mismatches 85; Indels 7; Gaps 3;

Qy 1 MDPEGGKRGSAENFWMGKSKKKEKKEKPTVTSTAMPYRYSNWLDRLVLMYCTMAAIIH 60
Db 1 MELEDLKGADNFRSMKMGKSKKKEKKEKPAVSIVTMFRIYAGWLDRLVLMYCTMAAIIH 60
Qy 61 GAALPLMLLVFGNMTDFANAG-ISRNTPTPIINESITNTQHFHNHLEEMTYTAYYY 119
Db 61 GVALPLMLLVFGNMTDFSFAGVGNVSKNST-----NMSEADKRAMFAK-LEEEMTYTAYYY 114
Qy 120 SGIGAGVLVAAIYQVSWFCLAAAGQILKIRKQFFHAIMRQIEGDFVDHVGELNRLTDD 179
Db 115 TGIGAGVLVAAIYQVSWFCLAAAGQILKIRKQFFHAIMRQIEGDFVDHVGELNRLTDD 174
Qy 180 VSKINEGIGDKVGMFFOSIATFTFTGTPGKWLTLVLAISPVGLSAAIWAAILSS 239
Db 175 VSKINEGIGDKVGMFFQAMATFFGGFIIGTRGKWLTLVLAISPVGLSAGIWAAILSS 234
Qy 240 FTDKELLAYAKAGAAVEVLAAIRTVIAFGGQKKELEKRYNNLEAKRIGIKKAITANIS 299
Db 235 FTDKELLAYAKAGAAVEVLAAIRTVIAFGGQKKELEKRYNNLEAKRIGIKKAITANIS 294
Qy 300 IGAAPLLIYAYALAFYVGSLSVLSSSEYTTGOVLTVFVSSVLIGAFSGQASPIEAFANA 359
Db 295 MGAAPLLIYAYALAFYVGSLSVLSSSEYTTGOVLTVFVSSVLIGAFSGQASPIEAFANA 354
Qy 360 RGAAYEFKIIDNKPISIDTSKSGHKPDNKGNEFNHVSFSPSRKEVKILKGLNLKVQ 419
Db 355 RGAAYEFKIIDNKPISIDTSKSGHKPDNKGNEFNHVSFSPSRKEVKILKGLNLKVQ 414
Qy 420 SGQTVLVGNSGCGSKSTTVQLMQRLDPTDGMVCDIGQDINTINVRHLREITGVVSGEPV 479
Db 415 SGQTVLVGNSGCGSKSTTVQLMQRLDPTDGMVCDIGQDINTINVRHLREITGVVSGEPV 474
Qy 480 LFATTIAENIRYGRNVTMDIEKAVKANNAYDFIMKLPNKFDPLVGERGARGSGGQQR 539
Db 475 LFATTIAENIRYGRNVTMDIEKAVKANNAYDFIMKLPNKFDPLVGERGARGSGGQQR 534
Qy 540 TAIARALVRNPKILLDLDEATSEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVI 599
Db 535 TAIARALVRNPKILLDLDEATSEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVI 594

Qy 600 AGFDGVIIVEGNHDELMKEKGIYFKLVMTQTRGNEIELENATGESKSESDALEMSPKDS 659
Db 595 AGFDGVIIVEGNHDELMKEKGIYFKLVMTQTRGNEIELENATGESKSESDALEMSPKDS 654
Qy 660 GSSILKRRSPRSTHAPOGODRKLGTKEDLNENVPVSFWRLKLNSTENPFWVVGIFCA 719
Db 655 GSSILKRRSPRSTHAPOGODRKLGTKEDLNENVPVSFWRLKLNSTENPFWVVGIFCA 714
Qy 720 IINGLOPAFASIIISRIIGTFRDEDETPKRNNSMFSVLVLGIIISFTIFFLQGGTFG 779
Db 715 IINGLOPAFASIIISRIIGTFRDEDETPKRNNSMFSVLVLGIIISFTIFFLQGGTFG 774
Qy 780 KAGEILTKRLRYMFRSMRLQDVSWFDDPKNTTCALTITRLANDAAQVKAIGSLAVITO 839
Db 775 KAGEILTKRLRYMFRSMRLQDVSWFDDPKNTTCALTITRLANDAAQVKAIGSLAVITO 834
Qy 840 NIANLGTGIIISLIYGWLTLLLAIVPIIAIAGVEMKMLSGQALKDKKKELEGAGKIAT 899
Db 835 NIANLGTGIIISLIYGWLTLLLAIVPIIAIAGVEMKMLSGQALKDKKKELEGAGKIAT 894
Qy 900 EAIENFRVTSITREQKPEYMAQSLQVPIYRNAMKKAHVFGITFEFFQAMMYFSYACFR 959
Db 895 EAIENFRVTSITREQKPEYMAQSLQVPIYRNAMKKAHVFGITFEFFQAMMYFSYACFR 954
Qy 960 FGAYLVANEFNFDQVLLVFSALVFGAMAVQVSSFPADYAKAVSAHVIMIEKSPLI 1019
Db 955 FGAYLVANEFNFDQVLLVFSALVFGAMAVQVSSFPADYAKAVSAHVIMIEKSPLI 1014
Qy 1020 DYSYPHGLKPNLTGCVNFTFNEVFNYPTRDPVQLGILSLEVKKGTQTLALVSGSGCGKST 1079
Db 1015 DYSYPHGLKPNLTGCVNFTFNEVFNYPTRDPVQLGILSLEVKKGTQTLALVSGSGCGKST 1074
Qy 1080 VVQLLEFYDPLAGSVLDGKEIKHLNVOVLAHIGVIVSQBPILFDCSIAENIAYGNSR 1139
Db 1075 VVQLLEFYDPLAGSVLDGKEIKHLNVOVLAHIGVIVSQBPILFDCSIAENIAYGNSR 1134
Qy 1140 VYSHEIMQAKEANHHFIETLPEKYNTRYGDKGTQSLSGGKQRIATARALVRQPHILL 1199
Db 1135 VYSHEIMQAKEANHHFIETLPEKYNTRYGDKGTQSLSGGKQRIATARALVRQPHILL 1194
Qy 1200 LDEATSAIDTSEKVVQVQALDKAREGRTCVIAHRLSTIQNALDILVVFQNKVKEHGHQ 1259
Db 1195 LDEATSAIDTSEKVVQVQALDKAREGRTCVIAHRLSTIQNALDILVVFQNKVKEHGHQ 1254
Qy 1260 QLLAOKGIYFMSVQAGAKR 1280
Db 1255 QLLAOKGIYFMSVQAGAKR 1275

RESULT 4

DWNSI
multidrug resistance protein 1 - mouse
N:Alternate names: P-glycoprotein 1
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: A33719; A25057; I57510
R:Raymond, M.; Gros, P.
A:Title: Mammalian multidrug-resistance gene: correlation of exon organization with s
A:Reference number: A33719; MUID:89367274
A:Accession: A33719
A:Molecule type: DNA
A:Residues: 1-1276 <RAY>
R:Gros, P.; Croop, J.; Housman, D.
Cell 47, 371-380, 1986
A:Title: Mammalian multidrug resistance gene: complete cDNA sequence indicates strong
A:Reference number: A25057; MUID:87028229
A:Accession: A25057
A:Molecule type: mRNA
A:Residues: 1-1276 <GRO>
A:Cross-references: GB:M14757; NID:g199100; PIDN:AAA79005.1; PID:g387426
R:Raymond, M.; Gros, P.
Mol. Cell. Biol. 10, 6036-6040, 1990

A:Title: Cell-specific activity of cis-acting regulatory elements in the promoter of the
A:Reference number: 157510; MUID:91042535
A:Accession: 157510
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-21 <RES>
A:Cross-references: GB:M60348; NID:g199102; PIDN:AAA39513.1; PID:g554199
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: mdrl (pgp1)
A:Introns: 21/2; 38/3; 92/1; 111/2; 175/2; 233/3; 274/2; 332/3; 370/3; 407/3; 449/3; 517
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane pr
F:1-637,653-1276/Region: duplication
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:550-554/Region: nucleotide-binding motif B
F:1051-1247/Domain: ATP-binding cassette homology <ABC2>
F:1068-1075/Region: nucleotide-binding motif A (P-loop)
F:1194-1198/Region: nucleotide-binding motif B
F:73,91,96,103/Binding site: carboxylate (Asn)
F:432/Binding site: ATP (Lys) #status predicted
F:1074/Binding site: ATP (Lys) #status predicted

Query Match 82.0%; Score 5304.5; DB 1; Length 1276;
Best Local Similarity 80.2%; Pred. No. 2.2e-302;
Matches 1027; Conservative 133; Mismatches 114; Indels 7; Gaps 4;

QY 1 MDPEGRKGSSEKFNWFKGSKSEKKEKPTVSTFAMFRYSNWLRLYLMLVGMALIIH 60
Db 1 MEFEENLGRADKNFKSMGKSKSEKKEKPAVGFGFMFYADWLDCILMILGTAAIIH 60
QY 61 GAALPLMLVFGNMTDFANAGISRNKTFPVILINESITNTQFIN-HLEEEMTYAYY 119
Db 61 GTLLPLMLVFGNMTDFTAEAS---ILPSINQSPNLTILISNSLSEEMAIYAYY 117
QY 120 SGIGAGVLVAYIOVSWCILAAGRLKIRKQFFHAIMRQIEGWFVDHYGELNTRLTDD 179
Db 118 TGIGAGVLVAYIOVSLWCLAAGRLKIRKQFFHAIMRQIEGWFVDHYGELNTRLTDD 177
QY 180 VSKINEGIDKVGKMGFFOSIATFTGFTVGPFGKTLVLILASPVLGLSAAIWAIIIS 239
Db 178 VSKINDIGIDKMGFFOSITFTLAGFTIGSKMGLTLVLILASPVLGLSAAIWAIIIS 237
QY 240 FTDKELLAYAKAGAVEEVLAAITRTVAFGGQKKELERYNKNLEAKRIGIKKAITANIS 299
Db 238 FTNKLQAYAKAGAVEEVLAAITRTVAFGGQKKELERYNKNLEAKRIGIKKAITANIS 297
QY 300 IGAAFLIYASYALATWYGTSLVLSSEYTTIGQVLTVFVSLIGAFSIGQASPSIEAFANA 359
Db 298 IGAYLLVYASYALATWYGTSLVLSSEYTTIGQVLTVFVSLIGAFSIGQASPSIEAFANA 357
QY 360 RGAAYEFKIDKPNKSDTSYKSGHPPDNILKGNLEKKNHFSPYSRKEVKILKGLNLKQV 419
Db 358 RGAAYEFKIDKPNKSDTSYKSGHPPDNILKGNLEKKNHFSPYSRKEVKILKGLNLKQV 417
QY 420 SGQTVLVGNSSGCKSTTVQMLQRLYDPTDMVCDIGQDITINVRHLREITGVWSQEPV 479
Db 418 SGQTVLVGNSSGCKSTTVQMLQRLYDPTDMVCDIGQDITINVRHLREITGVWSQEPV 477
QY 480 LFATTAAENIRYGRNVTWDEIEKAVKEANAYDFIMKLPNKFDTLYGERGARGSGQKQR 539
Db 478 LFATTAAENIRYGRNVTWDEIEKAVKEANAYDFIMKLPNKFDTLYGERGARGSGQKQR 537
QY 540 IATARALVRNPKLLLDDEATSDLTSEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVI 599
Db 538 IATARALVRNPKLLLDDEATSDLTSEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVI 597
QY 600 AGFDDGVIVKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGESKESDSEDALEMSPKDS 659
Db 598 AGFDDGVIVKGNHDELMKEKGIYFKLVMTQTRGNEIELENATGESKESDSEDALEMSPKDS 657

QY 660 GSSILIKRRSTERSIHQODRKLGTEDINENVPVSFWRILKLNSTENPYPVVGIFCA 719
Db 658 KSPULI-RRSIYRSVHRKQDQERRLSMKEAYDEDPVLSFWRILNLSNPYLLVGVLC 716
QY 720 IINGGLQAFASIIFSRIGIIFTRDEDEPKRQNSNMFSLFLVLGLIISFTTFFLQGTFFG 779
Db 717 VINGCIQPAIVFISRVIGVFSRDDHDEYKQNCNLFSLFLVGLISFTTFFLQGTFFG 776
QY 780 KAGEILTKRLRYMVRFSMLRQDVSDFDPKNTTGTALTRLANDAAQVKAIGSLRAVITQ 839
Db 777 KAGEILTTRVRYMVRFSMLRQDVSDFDPKNTTGTALTRLANDAAQVKAIGSLRAVITQ 836
QY 840 NIANLGTGIIISLYGWLTLALLAIVPIITAIAGVVMKMLSQALKDKELEGACKIAT 899
Db 837 NVANLGTGIVLSLYGWLTLALLAIVPIITAIAGVVMKMLSQALKDKELEGACKIAT 896
QY 900 EAIENFTVYSLTRQKFEFYMAQSLQVYPYRNSLRKAHIFGVFSFISITQAMMYFSYACFR 959
Db 897 EAIENFTVYSLTRQKFEFYMAQSLQVYPYRNSLRKAHIFGVFSFISITQAMMYFSYACFR 956
QY 960 FGAYLVANEFMNFQDVLVLSFALVFGAMAVGVSSPAPDYAKAKVSAAHVIMIEKSPLI 1019
Db 957 FGAYLVAAQLMTFENMLVFSVAVVFGAMAAGNTSSPAPDYAKAKVSAAHVIMIEKSPLI 1016
QY 1020 DSYSPHGLKPTLEGNTFENNVFNYPTRDIPVLOGLSLEVKKGOTLALVSSGCGKST 1079
Db 1017 DSYSTGLKPTLEGNTFENNVFNYPTRDIPVLOGLSLEVKKGOTLALVSSGCGKST 1076
QY 1080 VOLLERFYDPLAGSVLIDGKEIKHLNVQWLAHLGIVSQEPIFLDCSTAEINAYGDNRS 1139
Db 1077 VOLLERFYDPMAGSVFLDGEIKELQNLVQWLAHLGIVSQEPIFLDCSTAEINAYGDNRS 1136
QY 1140 VVSHEIMQAANKANHHIEFLPEKYNTRVGDKGTQSGGQKQRIATARALVRPHILL 1199
Db 1137 AVSHEIVRAAKENANTHQFIDSLPKYNTRVGDKGTQSGGQKQRIATARALVRPHILL 1196
QY 1200 LDEATSALETESKVVQEAIDKAREGRTCVIAHRLSTIONADLIIVFONGVKKEHGTHQ 1259
Db 1197 LDEATSALETESKVVQEAIDKAREGRTCVIAHRLSTIONADLIIVFONGVKKEHGTHQ 1256
QY 1260 QLLAOKGIYFMSIVQAGAKR 1280
Db 1257 QLLAOKGIYFSM-VQAGAKR 1275

RESULT 5
JH0502

p-glycoprotein - rat

N:Alternate names: multidrug resistance protein mdrlb

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Mar-1992 #sequence revision 31-Mar-1992 #text_change 02-Feb-2001

C:Accession: JH0502; S22353; S22352

R:Silverman, J.A.; Raulio, H.; Gant, T.W.; Thorgerirsson, S.S.

Gene 106, 229-236, 1991

A:Title: Cloning and characterization of a member of the rat multidrug resistance (mdr)

A:Reference number: JH0502; MUID:92039081

A:Accession: JH0502

A:Molecule type: mRNA

A:Residues: 1-1277 <SIL>

A:Cross-references: GB:M62425

R:Deuchars, K.L.; Duthie, M.; Ling, V.

Biochim. Biophys. Acta 1130, 157-165, 1992

A:Title: Identification of distinct p-glycoprotein gene sequences in rat.

A:Reference number: S22351; MUID:92223089

A:Accession: S22353

A:Molecule type: DNA

A:Residues: 1212-1226, '1', 1228-1277 <DEU>

A:Cross-references: EMBL:X61104; NID:g56890; PIDN:CAA43416.1; PID:g1334219

A:Accession: S22352

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1212-1226, '1', 1228-1270, 'SV', 1271-1277 <DE2>

A:Cross-references: EMBL:X61103; NID:g56888; PIDN:CAA43415.1; PID:g1334218

C;Comment: This protein has the nucleotide binding motifs and ATP binding active transpo
C;Genetics:
A;Gene: mdrlb
C;Superfamily: multidrug resistance protein; ATP-binding cassette homology
C;Keywords: ATP; duplication; glycoprotein; membrane protein; nucleotide binding; P-loop
F;109-603/Domain: ATP-binding cassette homology <ABC1>
F;426-434/Region: nucleotide-binding motif A (P-loop)
F;550-554/Region: nucleotide-binding motif B
F;1053-1248/Domain: ATP-binding cassette homology <ABC2>
F;1070-1078/Region: nucleotide-binding motif A (P-loop)
F;1195-1199/Region: nucleotide-binding motif B
F;432/Binding site: ATP (Lys) #status predicted
F;1076/Binding site: ATP (Lys) #status predicted

Query Match 81.0%; Score 5236; DB 2; Length 1277;
Best Local Similarity 80.1%; Pred. No. 2.3e-298;
Matches 1029; Conservative 132; Mismatches 110; Indels 14; Gaps 9;

QY 1 MDPEGGRKSAENFMWKGKSKKEKKPTVSTFAMFRYSNNLDRLYMLVGTMAAIIH 60
DB 1 MEFEELNGRADNFKSMGKSKKE-KEKAPAVIGFMFRYADWLKCLMALGTAAIIH 59
QY 61 GAALPLMLVFGNMNDSFANAGISRN-KTPPVVINESITNNTQHFIN--HLEEMTYAY 117
DB 60 GTLLPLMLVFGYMTDSFTP---SRDPHSDRAITNQSEINST-HTVSDTSLEEDMAMAY 115
QY 118 YYSIGAGVLVAAYIQVFSCLAAAGROILKIRQFPHAIMRQELGWFVDVHVGELNRLT 177
DB 116 YTTGIGAGVLTVAYIQVSLCLAAAGROIKIRQFPHAIMNQELGWFVDVNDAGELNRLT 175
QY 178 DVSKINEGIDGKVMFQSFQSTAFPTFTGIVGFTPGWKLTLVLAISPVLSAAIAWAKIL 237
DB 176 DVSKINDGIDGLGMFQSFITFSAGFIIGFISGWKLTLVLAISPVLSAAIAWAKVL 235
QY 238 SSTDKELLAYAKAGAAVEVLAAIRTVIAFGQKKELERYNNKLEAKRIGIKKAITAN 297
DB 236 TSFTNKELQAYAKAGAAVEVLAAIRTVIAFGQKKELERYNNKLEAKRIGIKKAITAN 295
QY 298 ISIGAEFLIYASALAFWGTSLVLSSEYTIQGLVTFVSLIGAFSIGOASPSTEAFA 357
DB 296 ISIGAILLVYASALAFWGTSLVLSSEYTIQGLVTFVSLIGAFSIGOASPSTEAFA 355
QY 358 NARGAAVEIFKIIDNPKSIDSYSGKSHKPNIKGNLFKNHFSYPSRKEVKILKGLNLK 417
DB 356 NARGAAVEIFKIIDNPKSIDSFSTGKHKPSIMGNLFKNHFSYPSRKEVKILKGLNLK 415
QY 418 VQSGQTVALVNSCGGKSTTVQLMORLYDPTDGMVCIDGQDIRTNVRLHREITGVVSQE 477
DB 416 VKSGQTVALVNSCGGKSTTVQLMORLYDPTDGMVCIDGQDIRTNVRLHREITGVVSQE 475
QY 478 PVLFTATTAEINIRYGRNVMTDEIKAVKEANAYDFIMKLPNKPDTLLVGERGARGSGQK 537
DB 476 PVLFTATTAEINIRYGRNVMTDEIKAVKEANAYDFIMKLPNKPDTLLVGERGARGSGQK 535
QY 538 QRIATARALVRNPKILLDEATSDALTESEAVQVALDKARKGRTTIVIAHRLSTVRNAD 597
DB 536 QRIATARALVRNPKILLDEATSDALTESEAVQVALDKARKGRTTIVIAHRLSTVRNAD 595
QY 598 VIAGFDGCVIVKGNHDELMKEGIFKFLVMTQTRGNEIELENATGESKESDALEMSPK 657
DB 596 VIAGFDGCVIVKGNHDELMKEGIFKFLVMTQTRGNEIELENATGESKESDALEMSPK 655
QY 658 DSGSGLIKRRSTRRSIHAPOQDRKLGTKEDLNENVPVSFWRILKLNSTWPFYVVGIF 717
DB 656 ESKSPLI-RRSIRRSIHRQDQERRLSKEDVDVPMVSFWQILKLNISEWPLVVGVL 714
QY 718 CAIINGLOPAPSIIFSRIIGFTRDEDPETHKQNSNMFVFLVLGLISFITEFLOGFT 777
DB 715 CAVINGCIPVFAIVFSKIVGVFSRDDHETKQRNCNLFSLFLVMGMSISVTVYFQGT 774
QY 778 FGKAGEILTKRLRYVWFSRLRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSR LAVI 837
DB 775 FGKAGEILTKRLRYVWFSRLRQDVSFDDHDKNTTGSTLTRLASDASNVKMGMSRLAVV 834

QY 838 TONTANLGTGIIIS--LIYQWQLTLLIIAIVPIIAIAGVWEMKMLSGOALKDKKLEBGAG 895
DB 835 TQNVANLGTGIIISLVLYVYGWQLTLLVWIPIVLVGLGIIEMKLLSGOALKDKKLEISG 894
QY 896 KIATEAIENFRTVSVLSTREQKFEFYMAQSLQVPTNLSRKAHIFGVFSFTQAMMYFSYA 955
DB 895 KIATEAIENFRTVSVLSTREQKFEFYMAQSLQVPTNLSRKAHIFGVFSFTQAMMYFSYA 954
QY 956 GCFRFGAYLVANFNFQDVLVLSAIVFGAMAVQVSSPAPDYAKAKVSAAHVIMIEK 1015
DB 955 ACFRFGAYLVARELMTFENVMLVFSVAVFGMAAGNTSSPAPDYAKAKVSAHIIIEK 1014
QY 1016 SPLIDSYSPHGLKPNLTLEGNTTFENVFNPTRPDIPVLOGLSLEVYKGGTALVVGSSGC 1075
DB 1015 IPEIDSYSTEGCLKPNWLGKNGKFNKFNKPNIPVLOGLSFEVKKGTQLVVGSSGC 1074
QY 1076 GKSTVVQLLRYDPDPLAGSVLIDGKEIKHLNVQWRAHLGIVSQEPILFDCSIAENTAYG 1135
DB 1075 GKSTVVQLLRYFNPDMAGTFLDGEIKQLNVQCVRA-LGIVSQEPILFDCSIAENTAYG 1133
QY 1136 DNSRVVSHEELMQAAKEANIIHFIETLPEKYNTRVGDGTOLSGGQKORIAIARALVRQP 1195
DB 1134 DNSRVVSHEELVRAAREANIIHQFIDSLPEKYNTRVGDGTOLSGGQKORIAIARALVRQP 1193
QY 1196 HILLDEATSDALTESEKVOEALDKAREGRTCIIVIAHRLSTIONADLIVVQNGKVKHE 1255
DB 1194 HILLDEATSDALTESEKVOEALDKAREGRTCVVIAHRLSTIONADLIVVQNGKVKHE 1253
QY 1256 GTHOQLLAQKGIYFSMISVQAGAKR 1280
DB 1254 GTHOQLLAQKGIYFSM--VQAGAKR 1276

RESULT 6
DWSIA
Multidrug resistance protein la - mouse (fragment)
N:Alternate names: P-glycoprotein la
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: A34175
R:Hsu, S. I. H.; Lothstein, L.; Horwitz, S. B.
J. Biol. Chem. 264; 12053-12062, 1989
A:Title: Differential overexpression of three mdr gene family members in multidrug-re
A:Reference number: A34175; MUID: 89308614
A:Accession: A34175
A:Molecule type: mRNA
A:Residues: 1-1104 <HSU>
A:Cross-references: GB:M24417; GB:J04839; NID:g200329; PIDN:AAA03243.1; PID:g200330
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c
tructurally and functionally unrelated lipophilic antitumor drugs.
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane
F;251-258/Domain: ATP-binding cassette homology <ABC1>
F;234-428/Domain: ATP-binding motif A (P-loop)
F;375-379/Region: nucleotide-binding motif B
F;877-1073/Domain: ATP-binding cassette homology <ABC2>
F;894-901/Region: nucleotide-binding motif A (P-loop)
F;1020-1024/Region: nucleotide-binding motif B
F;257/Binding site: ATP (Lys) #status predicted
F;900/Binding site: ATP (Lys) #status predicted

Query Match 77.0%; Score 4982; DB 1; Length 1104;
Best Local Similarity 88.9%; Pred. No. 1.4e-283;
Matches 981; Conservative 64; Mismatches 58; Indels 0; Gaps 0;

QY 178 DDVSKINEGIDGKVMFQSFQSIATFFTFIVGFTPGWKLTLVLAISPVLSAAIAWAKIL 237
DB 1 DDVSKINEGIDGKVMFQSIATFFTFIVGFTPGWKLTLVLAISPVLSAAIAWAKIL 60
QY 238 SSTDKELLAYAKAGAAVEVLAAIRTVIAFGQKKELERYNNKLEAKRIGIKKAITAN 297
DB 61 SSTDKELLAYAKAGAAVEVLAAIRTVIAFGQKKELERYNNKLEAKRIGIKKAITAN 120

```
QY 298 ISIGAAFLIIYASYALAFWYGTSLVLSSEYTIQOVLTWFFSVLIGAFSIGOASPSIEAFA 357
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 ISMGAAFLIIYASYALAFWYGTSLVLSSEYTIQOVLTWFFSVLIGAFSIGOASPSIEAFA 180
QY 358 NARGAAEYFIKIDNKPISIDSYSGHKKPNKIKGNLFHFSYPSKRKEVKILKGLNLK 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 NARGAAEYFVIKIDNKPISIDSYSGHKKPNKIQGNLFHFSYPSKRKEVKILKGLNLK 240
QY 418 VOSGOTVALVNSGCGKSTTVOLMORLYDPTDGMVCIQDQDIRTINVRHLREITGVVSQE 477
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 VKSGOTVALVNSGCGKSTTVOLMORLYDPLDGMVSIQDQDIRTINVRHLREITGVVSQE 300
QY 478 PVLFAATTIAENIRYGRENVMTDETEKAVKEANAYDFIMKLPNKFDTLVGBERGARLSGGOK 537
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 PVLFAATTIAENIRYGRENVMTDETEKAVKEANAYDFIMKLPNKFDTLVGBERGARLSGGOK 360
QY 538 ORIAIARALVNPKILLDEATSDALTESAVVOALDKARKGRTTIVIAHRLSTVRNAD 597
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 ORIAIARALVNPKILLDEATSDALTESAVVOALDKARKGRTTIVIAHRLSTVRNAD 420
QY 598 VIAGFDGCVIVEKGNHDELMEKGIYFKPLVMTQTRGNEIELENATGESKESDALEMSPK 657
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 VIAGFDGCVIVEKGNHDELMEKGIYFKPLVMTQTRGNEIELENATGESKESDALEMSPK 480
QY 658 DSGSSLIKRRSTRSIHAPQODRKLGTREDLNENVPVSFWRIKLKLNSTWPFYVVGIF 717
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 DSGSSLIKRRSTRSIHAPQODRKLGTREDLNENVPVSFWRIKLKLNSTWPFYVVGIF 540
QY 718 CAIINGLOPAFSTIIFRIGTREDPTTKQKNMSEVLVLGLIISFTIFFLQGGT 777
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 541 CAIINGLOPAFSTIIFRIGTREDPTTKQKNMSEVLVLGLIISFTIFFLQGGT 600
QY 778 FGKAGEILTRLRYMFRSLRODVSFDDPKNTTCALATRLANDAAQVKGALGSLAVI 837
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 601 FGKAGEILTRLRYMFRSLRODVSFDDPKNTTCALATRLANDAAQVKGALGSLAVI 660
QY 838 TONTANIGTGIIISLYGWOLTLALLAIPIIATAGVEMKMLSGQALDKKLESGAGKI 897
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 661 TONTANIGTGIIISLYGWOLTLALLAIPIIATAGVEMKMLSGQALDKKLESGAGKI 720
QY 898 ATEAENFRVTSITREKFEYMAQSLQVPPYNSLRKAHITGVSFSTQAMMYFSYAGC 957
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 721 ATEAENFRVTSITREKFEYMAQSLQVPPYNSLRKAHITGVSFSTQAMMYFSYAGC 780
QY 958 FRFGAYLVANEFMNFODVLLVFSIAIVFGAMAYGOVSSFPADYAKAKVSAAHVIMIIKSP 1017
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 781 FRFGAYLVANEFMNFODVLLVFSIAIVFGAMAYGOVSSFPADYAKAKVSAAHVIMIIKSP 840
QY 1018 LIDSYPHGLKPNLEGNVTNEVFNYPTRPDIPLVQGLSLEVKKGTALVGSNGCGK 1077
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 841 EIDSYSTQGLKPNMLEGNVQFSGVFNYPTRPSIPVQGLSLEVKKGTALVGSNGCGK 900
QY 1078 STVQVLLERFYDPLAGSLVDKGEIKHLNVMQVLAHLGVISOEPLFPCSTAEANIAYGDN 1137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 901 STVQVLLERFYDPLAGSLVDKGEIKHLNVMQVLAHLGVISOEPLFPCSTAEANIAYGDN 960
QY 1138 SRVVSHEEIMQAAKEANIHFIETLPEKYNTRVGDKGTQVLSGQKQRTAIARALVRPHI 1197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 961 SRVVSHEEIMQAAKEANIHFIETLPEKYNTRVGDKGTQVLSGQKQRTAIARALVRPHI 1020
QY 1198 LLLDEATSDALTESKVVQVQALDKAREGTCITVIAHRLSTTONADLIIVVFONGVKVKEHGT 1257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1021 LLLDEATSDALTESKVVQVQALDKAREGTCITVIAHRLSTTONADLIIVVFONGVKVKEHGT 1080
QY 1258 HQOLLAQKGIYFSMISVQAGAKR 1280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1081 HQOLLAQKGIYFSMISVQAGAKR 1103
RESULT 7
DVH03
multidrug resistance protein 3 - human
```

```
N:Alternate names: P-glycoprotein MDR3
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: J50051; S01346; A42213; I38015
R:van der Bliek, A.M.; Koolman, P.M.; Schneider, C.; Borst, P.
Gene 71, 401-411, 1998
A:Title: Sequence of mdr3 cDNA encoding a human P-glycoprotein.
A:Reference number: J50051; MUID:89138016
A:Accession: J50051
A:Molecule type: mRNA
A:Residues: 1-1279 <VA1>
R:van der Bliek, A.M.; Baas, F.; ten Houte de Lange, T.; Koolman, P.M.; van der Velde
EMBO J. 6, 3325-3331, 1987
A:Title: The human mdr3 gene encodes a novel P-glycoprotein homologue and gives rise
A:Reference number: S01346; MUID:8811519
A:Accession: S01346
A:Molecule type: mRNA
A:Residues: 856-1093 'FYDFGFO', 1094-1279 <VA2>
A:Cross-references: EMBL:X06181; NID:g34524; PIDN:CAA29547.1; PID:g34525
R:Lincke, C.R.; Smit, J.J.M.; van der Velde-Koerts, T.; Borst, P.
J. Biol. Chem. 266, 5303-5310, 1991
A:Title: Structure of the human MDR3 gene and physical mapping of the human MDR locus
A:Reference number: A42213; MUID:91161629
A:Accession: A42213
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1279 <LIN>
R:Smit, J.J.; Mol, C.A.; van Deemter, L.; Wagenaar, E.; Schinkel, A.H.; Borst, P.
Biochim. Biophys. Acta 1261, 44-56, 1995
A:Title: Characterization of the promoter region of the human MDR3 P-glycoprotein gen
A:Reference number: I38015; MUID:95200972
A:Accession: I38015
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-61, 'RGSSRVDLQAC' <RES>
A:Cross-references: EMBL:235284; NID:g1006662; PIDN:CA84542.1; PID:g1006663
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c
tructurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: GDB:PGY3; MDR3
A:Cross-references: GDB:120713; OMIM:171060
A:Map position: 7q21-7q21
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane
F:1-640,653-1279/Region: duplication
F:412-606/Domain: ATP-binding cassette homology <ABC1>
F:429-436/Region: nucleotide-binding motif A (P-loop)
F:553-557/Region: nucleotide-binding motif B
F:638-694/Domain: linker <LINK>
F:1052-1248/Domain: ATP-binding cassette homology <ABC2>
F:1069-1076/Region: nucleotide-binding motif A (P-loop)
F:1195-1199/Region: nucleotide-binding motif B
F:91,97/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:435/Binding site: ATP (Lys) #status predicted
F:1075/Binding site: ATP (Lys) #status predicted
Query Match 75.7% Score 4898; DB 1; Length 1279;
Best Local Similarity 75.4%; Pred. No. 1.4e-278;
Matches 971; Conservative 134; Mismatches 163; Indels 20; Gaps 9;
QY 1 MDPEGGRKG-----SAEKNFWKMGKKKKKKKPT--VSTFAMFYSNWLDRLYMLV 52
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MDLEAKNKTAWRPISAEQDF-ELGSSQKKRKKTKTVKMGVLTFLFRYSDQKLFMSL 59
QY 53 GTMAAIIHGAALPLMLLVFGNNMDSFANAGISRNKTFPVIINESITNNQHFHLEEM 112
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 GTIMAHGSGGLPLMMIVFEGMTDKFVDT--AGNFSFPVNFSLSLNPGK----ILLEEM 113
QY 113 TTYAYYISGIGAGVLAAYIQVSWFCLAAAGQILKIRKQFPHAIMROEIGNFVDHVGEL 172
| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: ||
Db 114 TRYAYYISGLGAGVLAAYIQVSWFLLAAGRQIRKQKFFHALLRQEIFGWFINDTTEL 173
```

QY	173	NTRLTDDVSKINIGIDKVMGFQSIATFTGFIIVGFTPGWKLTLVLAIISPVLGSAAI	232
Db	174	NTRLTDDISKISGIDKVMGFQAVATFAGFIVGFIIRGWKLTIVMAISPILGSAAV	233
QY	233	WAKILSFTDKELLAYAKAGAAVEVLAIRTVIATGQKKELERYKNKLEAKRIGIKK	292
Db	234	WAKILSAFSDKELAAAYAKAGAAEALGAIRTVIATGQKKELERYOKKLENAKEIGIKK	293
QY	293	AITANTISGAFLIIVASALAFWYGTSLVLSSEYITIGQVLTFFSVLIGAFSIGQASPS	352
Db	294	AISANTISMGTAFLIIVASALAFWYGTSLVLSKEYITIGNAMTVFFSILIGAFSVGQAAPC	353
QY	353	IEAFANARGAAEYFIKIIDNKPSIDSYSGKHGHPDNIKNLEKVNHFSPSRKEVKILK	412
Db	354	IDAFANARGAAVIFDIIDNPKIDSFSRGRKPDISKGNLENDVHFSPSRANVKILK	413
QY	413	GLNLKVGSGTVALVNGSGCKSTTVOLMORLYDPTDGMVICDQDIRTVNRLHREITG	472
Db	414	GLNLKVGSGTVALVNGSGCKSTTVOLIQRLYDPDEGTINIDQDIRNFNVNLYREIIG	473
QY	473	VVSQEPVLFTATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLNKFDTLVGERGAL	532
Db	474	VVSQEPVLFTTIAENICYCGNVMTDEIKKAVKEANAYEFIMKLPQKFDTLVGERGAOL	533
QY	533	SGGOKRIATARALVRNPKILLDEATSDLTSEAVVOVALDKARKGRTTIVIAHRLST	592
Db	534	SGGOKRIATARALVRNPKILLDEATSDLTSEAEVQALDKAREGRTTIVIAHRLST	593
QY	593	VRNADVTAGFDGIVVEKGNHDELMKEGYFKLVMTQTRNETELEENATGESKSDAL	652
Db	594	VRNADVTAGFDGIVVEGSHSELKKEGYFKLVNMQTSSQIQSEEF--ELNDEKAAT	651
QY	653	EMSPKDSGSLIKRRSTRSHIAPQODRKLTGKED--LNVNVSYSFWRILKLNSTWEPY	711
Db	652	RMAPNGKSRLF--RHSTQKNLQMSQCKSLDVEDTGLEANVPVPSFLVKLKLNTWEPY	710
QY	712	FVVGIFCAINGGLQPAFSIIFSRIGITFTRDEPTEKRONSNMFSVFLVLGLIISITF	771
Db	711	FVVGIVCAIANGGLQPAFSVIFSEIIAIFGPGDD--AVKQKCNFSLIFLFLGIISFTF	769
QY	772	FLOQFTFKAGEIILTKRLRYMVFRSMURQDVSWFDDPKNTTGALTTRIANDAAQVKAIG	831
Db	770	FLOQFTFKAGEILTRRLRSMFAKMLRQDMSWFDHKNSTGALSTLATDAAQVOGATG	829
QY	832	SLAVITONIANLNGIILISILYQWLTLLLAIVPIIATAGVYEMKMLSGOALKDKKEL	891
Db	830	TLALIAQNIANLNGIILISILYQWLTLLLAIVPIIATAGVYEMKMLLAGNAKDKKEL	889
QY	892	EGAGKIATEAIENTRTVSLTREQKFEYMYAQSLQVYPYRNSLRKAHIFGVSFSTIQAMMY	951
Db	890	EAAGKIATEAIENTRTVSLTRQERKFESMYVEKLYGPPYRNSVQKAHIYGITFSIQAFMY	949
QY	952	FSYACGFRGAYLVANEPNFQDVLIVFSAIVFGAMAVQVSSFAPDYAKAKVSAAHVIM	1011
Db	950	FSYACGFRGAYLVYNGHMRFDVILVFSIVFGAVALGHASSFAPDYAKAKLSAHLFUM	1009
QY	1012	ITEKSPIDISYSPHGLKNTNLEGNVFNENYPTRPDIPIVLOGLSLVEKKGQTLALVG	1071
Db	1010	LFERPLDISYSEELKFKDPKEGNTFNEVFNYPTRANVPVLOGLSLVEKKGQTLALVG	1069
QY	1072	SSGCGKSTVVQLLRYFDPLAGSVLIDGKETKHLNVOMLRAHLGIVSOEPIIFDCSTAEH	1131
Db	1070	SSGCGKSTVVQLLRYFDPLAGTVLLDQGEAKLNVOWLRAQLGIVSOEPIIFDCSTAEH	1129
QY	1132	IAYGNSRVVSHETMQAKENIHHFETLPEKYNTRVGDGTLQSGGOKRIATARAL	1191
Db	1130	IAYGNSRVVSDIEVSAKAAANIHPFIETLPHKYETRVGDGTLQSGGOKRIATARAL	1189
QY	1192	VROPHTLLDEATSDLTSEKVVQEQALDKAREGRTCTIVIAHRLSTIONADLIIVVFQNG	1251
Db	1190	IROPQILLDEATSDLTSEKVVQEQALDKAREGRTCTIVIAHRLSTIONADLIIVVFQNG	1249
QY	1252	VKEHGTHQOULLAOKGIYFSMVSVQAGAK	1279

Db	1250	VKEHGTHQOULLAOKGIYFSMVSVQAGTQ	1277
RESULT 8			
DVMS2			
N:multidrug resistance protein 2 - mouse			
N:Alternates names: p-glycoprotein MDR2			
C:Species: Mus musculus (house mouse)			
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001			
C:Accession: A30409; S70711			
R:Gros, P.; Raymond, M.; Bell, J.; Housman, D.			
Mol. Cell. Biol. 8, 2770-2778, 1988			
A:Title: Cloning and characterization of a second member of the mouse mdr gene family			
A:Reference number: A30409; MUID:88302195			
A:Accession: A30409			
A:Molecule type: mRNA			
A:Residues: 1-1276 <HSU>			
A:Cross-references: GB:J03398; NID:g199109; PIDN:AAA39516.1; PID:g387428			
R:Kirschner, L.S.			
Nucleic Acids Res. 24, 2829-2834, 1996			
A:Title: De novo generation of simple sequence during gene amplification.			
A:Reference number: S70711; MUID:96313253			
A:Accession: S70711			
A>Status: nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residues: 43-92 <KIR>			
A:Cross-references: EMBL:U46839; NID:g128142; PIDN:AAC52722.1; PID:g128143			
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1996			
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c			
tructurally and functionally unrelated lipophilic antitumor drugs.			
C:Genetics:			
A:Gene: mdr2			
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology			
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; p-loop; transmembrane			
F:409-603/Domain: ATP-binding cassette homology <ABC1>			
F:426-433/Region: nucleotide-binding motif A (P-loop)			
F:550-554/Region: nucleotide-binding motif B			
F:1049-1245/Domain: ATP-binding cassette homology <ABC2>			
F:1067-1074/Region: nucleotide-binding motif A (P-loop)			
F:1192-1196/Region: nucleotide-binding motif B			
F:88,94/Binding site: nucleotide-binding (Asn) (covalent) #status predicted			
F:432/Binding site: ATP (Lys) #status predicted			
F:1072/Binding site: ATP (Lys) #status predicted			
Query Match	75.0%	Score 4850.5; DB 1; Length 1276;	
Best Local Similarity	74.6%;	Pred. No. 8.4e-276;	
Matches	958; Conservative 140; Mismatches 171; Indels 15; Gaps 8;		
QY	1	MDPEGKGRKSAEK--NFWKMGKSK--KEKKEKPTVSTFAMFRYSNWLDRMLVGTMA	56
Db	1	MDLEARNGTARLDGDFELGSLSNQGRKKKKVNLIGLLTLFRYSDWDQDKLFMFLGTLM	60
QY	57	AIITHGAALPLMLVFGNMTDSFA--NAGISRNKTFPVIINESTINNTQHFINHLEEDMTTY	115
Db	61	AIAHGSLPLMLVFGMTDKFDVNTG---NFSLPVNFSLSMLNPGK---ILEEMTRY	113
QY	116	AYTYSIGAGVVAAYIQVSWFCLAAAGRLTKRKOFFFHAIMRQETGWDFVDVHVGELNTR	175
Db	114	AYYSGLGGVLAAYIQVSWFCLAAAGRLTKRKOFFFHAIMRQETGWDFVDVHVGELNTR	173
QY	176	LTDDVSKINREGIDKVMGFQSIATFTGFIIVGFTPGWKLTLVLAIISPVLGSAAIWAK	235
Db	174	LTDDVSKISGIDKVMGFQAVATFAGFIVGFIIRGWKLTIVMAISPILGSLTAVWAK	233
QY	236	ILSFTDKELLAYAKAGAAVEVLAIRTVIATGQKKELERYKNKLEAKRIGIKKAIT	295
Db	234	ILSFTDKELAAAYAKAGAAEAPGAIRTVIATGQKKELERYOKKLENAKIKIKAIS	293
QY	296	ANISGAFLIIVASALAFWYGTSLVLSSEYITIGQVLTFFSVLIGAFSIGQASPSIEA	355
Db	294	ANISWGAFLIIVASALAFWYGTSLVLSKEYITIGNAMTVFFSILIGAFSVGQAAPCIDA	353

```
QY 356 FANARAAAEIIFKIIDNPSIDSKSGHKPDNIKGNLEKPNVHFSYSPSRKEVKILKGLN 415
Db 354 FANARAAAVIFDIIDNPNKIDSFGRHKPDNIKGNLESDVHFSYSPSRANIKILKGLN 413
QY 416 LKVGSGQTVALVNGSGCKSTTVQLMORLYDPTDGMVCIDGQDITINVRHLREITGVVS 475
Db 414 LKVGSGQTVALVNGSGCKSTTVQLLORLYDPTGKISIDQDILRNVRCLREIIGVVS 473
QY 476 QBPVLFTTAENIRYGRNVTDEIEKAYKEANAYDFIMKLPNKFTLVGERGARS 535
Db 474 QBPVLFTTAENIRYGRNVTDEIEKAYKEANAYDFIMKLPNKFTLVGERGARS 533
QY 536 QKQRIATARALVRNPKILLDEATSDLTSEAVVQVVALDKARKGRTTIVIAHRLSTVRN 595
Db 534 QKQRIATARALVRNPKILLDEATSDLTSEAVVQVVALDKARKGRTTIVIAHRLSTVRN 593
QY 596 ADVIAGFDGVIYVKGNDHLMKEKGIYFKLVMTQTRGNEIELENATGESKSDALEMS 655
Db 594 ADVIAGFDGVIYVQGSHELMKEGTYFRLVNMOTAGSQILSEFEVEVLSDEKAAGDVA 653
QY 656 PKDSSSLIKRRSTRRSIHAPQODRKLGTKEIDLNNVPSVSWRILKLNSTWPFYVVG 715
Db 654 P-NGWKARIFNSTKSLKSPH-ONRDUDEETNELDANVPVSLKVLKLNKTEWPFYVVG 711
QY 716 IFCAIINGLOPAPFSIIIFSRIGITFTRDEDPETKRONSMFSLVFLVLGIIITPFLOG 775
Db 712 TVCAIANGALQAPFSIIILSEMIAFPGDD-AVKQCKNMFSLVFLGLGVSEFFTFLOG 770
QY 776 PTFCKAGEILTKRLRYMVFMSRLRQDVSWFDDPKNTGTGALTTRLANDAAQVKAIGSRUA 835
Db 771 PTFCKAGEILTKRLSRAFAKRLQDMWFDHKNSTGALSTRLATDAAQVQCATGKLA 830
QY 836 VITQNIANLGTGIIISLYGQWLTLLLATVPIIATAGVVMKLSQALKDKKELEGAG 895
Db 831 LIAQNTANLGTGIIISLYGQWLTLLSVVPIAVAGIVEMKLAGNAKRDCKEMEAG 890
QY 896 KIAPEATENRTVVSILTQERKFESMVYKELHGPYRNSVRKAHIYGITFSIQAFMYFSA 955
Db 891 KIAPEATENRTVVSILTQERKFESMVYKELHGPYRNSVRKAHIYGITFSIQAFMYFSA 950
QY 956 GCRPFGAYLVANEFMFODVLLVFSALVYAGAMAVQVSSFADYAKAKVSAAHVIMITEK 1015
Db 951 GCRPFGSILVNGHMRKFDVILVFSALVGLGALCHASSFADYAKAKLSAAAYLSLEFR 1010
QY 1016 SPLTDSYSPHGLKPNTELEGNTNEVYFNYPTRPDIPVLQGLSLEKKGOTLALVSSGC 1075
Db 1011 QPLIDSYSGBGLPDRFEGSVTFNEVFNYPTRANVPVLQGLSLEKKGOTLALVSSGC 1070
QY 1076 GKSTVVOLLERFYDPLAGSVLIDKEIKHLNVOWLRAHLGIYVQBPIFDCSIAENIAYG 1135
Db 1071 GKSTVVOLLERFYDPLAGSVLIDQBAKKNVOWLRAQLGIYVQBPIFDCSIAENIAYG 1130
QY 1136 DNRVYVSHETMOAAKEANTHIEFTLPEKYNTRVGDGTQLSGQKQRIATARALVRQP 1195
Db 1131 DNRVYVPHDEIVRAAKENIHFIEFTLPKYNTRVGDGTQLSGQKQRIATARALVRQP 1190
QY 1196 HILLDEATSDLTSEKVVQVQALDKAREGRCTIVIAHRLSTIONADLIVVFONGKVKKEH 1255
Db 1191 RVLLDEATSDLTSEKVVQVQALDKAREGRCTIVIAHRLSTIONADLIVVIENGKVKKEH 1250
QY 1256 GTHQQLLAOKGIYFMSVQAGK 1279
Db 1251 GTHQQLLAOKGIYFMSVNIQAGTQ 1274
```

RESULT 9
I48123

p-glycoprotein isoform III - Chinese hamster

C:Species: Cricetulus griseus (Chinese hamster)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001

C:Accession: I48123

R:Endicott, J.A.; Sarangi, F.; Ling, V.

DNA Seq. 2, 89-101, 1991

A:Title: Complete cDNA sequences encoding the Chinese hamster P-glycoprotein gene fam
A:Reference number: I48121; MUID:921135896

A:Accession: I48123

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-1281 <RES>

A:Cross-references: GB:M60042; NID:g191168; PIDN:AAA6885.1; PID:g191169

C:Superfamily: multidrug resistance protein; ATP-binding cassette homology

C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop

F:412-606/Domain: ATP-binding cassette homology <ABC>

F:429-436/Region: nucleotide-binding motif A (P-loop)

F:1054-1250/Domain: ATP-binding cassette homology <ABC2>

F:1071-1078/Region: nucleotide-binding motif A (P-loop)

Query Match 74.4%; Score 4810; DB 2; Length 1281;

Best Local Similarity 73.5%; Pred. No. 2e-273;

Matches 947; Conservative 148; Mismatches 175; Indels 18; Gaps 8;

QY 1 MDPEGRKGSAEK-----NFWKMKKSK--KEKKEKPTVSTFAMPYRYSNWLDRLYMVG 53

Db 1 MDLEAARNGTARRPGTVEGDFELGSIISNQGRNKKKYNLIGLPLTFRYSDWQDKLPMILG 60

QY 54 TMAAIIHGAALPLMLLVFGNMTDSFA--NAGISRNKTFPPVIINESITNNTQHFNLHLEEM 112

Db 61 TMAIAHSGSLPLMMIVFGEMTKFVNNAG---NFSLPVNFSLUSMINPGR----ILEEM 113

QY 113 TTYAYYSGIGAGVLYAAYIQVSFWCLAGROILKIRKOFFHAIMRQETGWDVHDVGBL 172

Db 114 TRAYAYYSGLGGVLYAAYIQVSFWTLAAGROIKKIRQNFHAILRQEMGWFIDKGTTEL 173

QY 173 NTRLTDDVSKINEGIGDKVCMFQSTATPFTGTFVGTGKLTIVILAIISPVLGSAAI 232

Db 174 NTRLTDDISIKSEIGDKVGMFFQVATFAGFVIRGWKLTIVAIMAISPLGLSAAV 233

QY 233 WAKILSFTDKELLAYAKAGAVAEVLAARTVIAFGGOKKELERKNLEAKRIGTKK 292

Db 234 WAKILSFTSKELAAAYAKAGAVAEALGARTVIAFGGONKELERKQKHLNNAKIGTKK 293

QY 293 AITANISGAFLIIVASYALAFWYGTSLVLSSEXTIGQVLTVFFSVLIGAFSIGQASP 352

Db 294 AISANISMGTAFLIIVASYALAFWYGTSLVLSKEYTIGNATVFFSILIGAFSVGOAAPC 353

QY 353 IEAFANARGAAVEYFKIIDNPKSIDSKSGHKPDNIKGNLEKPNVHFSYSPSRKEVKILK 412

Db 354 IDAFANARGAAVYIFDIIDNPNKIDSFGRHKPDNIKGNLEKPNVHFSYSPSRANIKILK 413

QY 413 GLNLKVGSGQTVALVNGSGCKSTTVQLMORLYDPTDGMVCIDGQDITINVRHLREITG 472

Db 414 GLNLKVGSGQTVALVNGSGCKSTTVQLLORLYDPTGKISIDGQDILRNVRCLREIIG 473

QY 473 VVSQEPVLFTTAENIRYGRNVTMDTEKAVKEANAYDFIMKLPNKFTLVGERGARL 532

Db 474 VVSQEPVLFTTAENIRYGRNVTMEEIKAVKEANAYDFIMKLPNKFTLVGERGAQL 533

QY 533 SGQKQRIATARALVRNPKILLDEATSDLTSEAVVQVVALDKARKGRTTIVIAHRLST 592

Db 534 SGQKQRIATARALVRNPKILLDEATSDLTSEAEVQVVALDKARKGRTTIVIAHRLST 593

QY 593 VRNADVIAGFDGVIYVKGNDHLMKEKGIYFKLVMTQTRGNEIELENATGESKSDAL 652

Db 594 VRNADVIAGFDGVIYVQGSHELMKEGTYFRLVNMOTAGSQILSQEFVEVLSDEKAAD 653

QY 653 EMSPKDSSSLIKRRSTRRSIHAPQODRKLGTKEIDLNNVPSVSWRILKLNSTWEMP 711

Db 654 GMP-NGWKARIFNSTKSLKSPH-ONRDUDEETNELDANVPVSLKVLKLNKTEWEMP 712

QY 712 FVVGIFCAIINGLOPAPFSIIIFSRIGITFTRDEDPETKRONSMFSLVFLVLGIIISITP 771

Db 713 FVVGITVCAIINGALQAPFSIIILSEMIAFPGDD-AVKQCKNMFSLVFLGLGVSEFFTF 771

QY 772 FLQGTFTGKAGEILTKRLRYMVFMSRLRQDVSWFDDPKNTGTGALTTRLANDAAQVKAIG 831

Db	772	FLQFTFGKAGEILLTRLRSMFAKAMLRQDMSWFDYKNTSGALSTRLATDRAQVOGATG	831
QY	832	SRNAVITQIANLGTGIIISLIYQWLTLLLAIVPIIAIAGVEMKMLSGQALKKKEL	891
Db	832	TRLALIAQNTANLGTGIIISLIYQWLTLLLSVVPFIAVSGIVEMKMLAGNAKRDKKAL	891
QY	892	EGAGKIATEAIENFRVVSILTREQFEYMYAQSLOVPYRNSLRKAHIFGVFSITQAMMY	951
Db	892	EAAGKIATEAIENIRTVSILTQERKFESYVEKLHEPYRNSVQMAHIYGITISQAFMY	951
QY	952	FSVAGCFRGAYLVANEENFODVLVFSIAIVFGAMAVQVSSFAPDYAKAKVSAAHVIM	1011
Db	952	FSVAGCFRGAYLIVNGHMRFRDVLVFSIAIVFGAVALGHASSFAPDYAKAKLSAAHLFS	1011
QY	1012	IIEKSLPIDSYPHGLKPNLTLEGNVTNFVFNYPTRPDIPVLQGLSLEVKKGQIALVG	1071
Db	1012	LFERQPLIDSYSCEGLWPDKFEGSVTFNEVFNYPTRANPVLQGLSLEVKKGQIALVG	1071
QY	1072	SSCGKSTVVQLLERYFDPLAGSVLDGKEIKHLNVQWLRHLGIVSQRPILFDCSIAEN	1131
Db	1072	SSCGKSTVVQLLERYFDPMAGTVLDDGQAKKLNIQWLRHLGIVSQRPILFDCSIAEN	1131
QY	1132	IAYGDSRVVSHEEIMOAAKEANIHFIEIETPEKYNTRVGDGKTQSGGQKORIAARAL	1191
Db	1132	IAYGDSRVVSODEIVRAAKAANIHFIEIETLPQKYKTRVGDGKTQSGGQKORIAARAL	1191
QY	1192	VROPHILLDEATSALDTESEKVVQEQALDKAREGRCTIVIAHRLSTIONADLLVVFQNGK	1251
Db	1192	TROPVLLDEATSALDTESEKVVQEQALDKAREGRCTIVIAHRLSTIONADLLVVFQNGK	1251
QY	1252	VKEGTHQQLLAQKGIYFMSVQAGAK	1279
Db	1252	VKEGTHQQLLAQKGIYFMSVNIQAGAQ	1279
RESULT 10			
S41646			
P-glycoprotein - rat			
C:Species: Rattus norvegicus (Norway rat)			
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 02-Feb-2001			
C:Accession: S41646; S22354			
R:Brown, P.C.; Thorngairsson, S.S.; Silverman, J.A.			
Nucleic Acids Res. 21, 3885-3891, 1993			
A:Title: Cloning and regulation of the rat mdr2 gene.			
A:Reference number: S41646; MUID:93376516			
A:Accession: S41646			
A>Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 1-1278 <BRO>			
A:Cross-references: EMBL:L15079; NID:g310192; PIDN:AAA02937.1; PID:g310193			
R:Deuchars, K.L.; Duthie, M.; Ling, V.			
Biochim. Biophys. Acta 1130, 157-165, 1992			
A:Title: Identification of distinct P-glycoprotein gene sequences in rat.			
A:Reference number: S22351; MUID:92223089			
A:Accession: S22354			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1211-1278 <DEU>			
A:Cross-references: EMBL:X61105; NID:g56892; PIDN:CAA43417.1; PID:g1334220			
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology			
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop			
F:409-603/Domain: ATP-binding cassette homology <ABC1>			
F:426-433/Region: nucleotide-binding motif A (P-loop)			
F:1051-1247/Domain: ATP-binding cassette homology <ABC2>			
F:1068-1075/Region: nucleotide-binding motif A (P-loop)			
Query Match 74.28; Score 4796.5; DB 2; Length 1278;			
Best Local Similarity 73.7%; Pred. No. 1.2e-272;			
Matches 947; Conservative 146; Mismatches 177; Indels 15; Gaps 8;			
QY	1	MDPEGRKGSAREK--NFWKMGKS--KKEKKEKPTVSTFAMFRYSNNWLDRLYLVTGMA	56
Db	1	MDLEARNGTARRLDGDFELGSLISNQSRKKKKVNLNGLPLFRYSQWQDKLFLMLGTAM	60

QY	57	AIHGAALPLMVLFGNMTSFA-NAGISRNKTPPVIINESITNTQHFHINHLEEMTTY	115
Db	61	AIAGSGPLPMIVFGEMTKFDVNDAG--NFSLPVNFSLMNLNPGR---ILEEMTRY	113
QY	116	AYYSIGAGVLYAAYIQVSWCLAAQRILKTRQFFHAIHQEIGWDFVDHVGELNTR	175
Db	114	AYYSIGLGGVLLAAYIQVSWTAAQRIRKIROKFFHAILROEMGFQDKGTTELNTR	173
QY	176	LTDDVSKINEGIGDKVGMFFOSIATEFTGFTVGTQWKLTVLAIASPLVLSAAIWAQ	235
Db	174	LTDDISKISEGIGDKVGMFFQAIATFFAGFIVGFRGKLLTVLMAITAILGLSTAWAK	233
QY	236	ILSFTDKELLAYAKAGAAVEVLAAITRTVAFGGQKKELERYNKNLEAKRIGIKAIT	295
Db	234	ILSFTSKELAAAYAKAGAAVEALGAITRTVAFGGQKKELERYQKHLNKKIGIKAIS	293
QY	296	ANISGAFLIYASYALAFWYGTSLVLSSEYTTGOVLTFFSVLIGAFSGQASPSIEA	355
Db	294	ANISMGIAFLIYASYALAFWYGTSLVLSSEYTTIGNAMTFFSVLIGAFSGQAAPCIDA	353
QY	356	FANARGAAVEIFKIIDNKPSIDSYSGKHGPDNKGLEFKNVHFSPSRKEVKILKGLN	415
Db	354	FPNARGAAVIFDIIDNNPKIDSFSEGHKPDGSIKGNLEFSDVHFSPSRANIKILKGLN	413
QY	416	LKVSQGTVALVNGSGCGKSTTVQLMQRLYDPTDGMVCIQDQDITINVRHLREITGVV	475
Db	414	LKVSQGTVALVNGSGCGKSTTVQLLQRLYDPTDGTISIDQDINRNVRCLEFIVGVV	473
QY	476	QEPVLFAITIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGARISGG	535
Db	474	QEPVLFTTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPQKFDTLVGERGAQLSGG	533
QY	536	OKORIAARALVRNPKILLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRN	595
Db	534	OKORIAARALVRNPKILLDEATSALDTESEAVQAAALDKARKGRTTIVIAHRLSTVRN	593
QY	596	ADVIAGFDGVIKGNHDELMKEGIYFKLVMTQTRNEIELENATGESKESDALEMS	655
Db	594	ADVIAGFDGVIKGNHDELMKEGIYFKLVNMTQSGILSEEFVELSDEKAAGVA	653
QY	656	PKDSGSLIKRSTRHSIHAPOGQDKLCTK-EDLENVNPVSFWRILKLNSTWPFYV	714
Db	654	P-NGWKARIFRNSTKSLSSRAHQNRDLVDNELDANVPVSFLKVLRLNKTWPFYV	712
QY	715	GIFCAITINGLQAPAFSIIFSRIGITRDEDPETKQNSMFSVLVLGIISFITFFLQ	774
Db	713	GTLCALANGALQAPAFSIILSEMIATFPGDD-TVQKQCNMFSLVFLGLGVHSFFTFLLQ	771
QY	775	GFTFGKAGEILLKRLRYMFRSMRLQDVSDFDDPKNTTGALTTRLANDAAQVKAIGSRL	834
Db	772	GFTFGKAGEILLKRLSMFAKMLRQDMSWFDHKNSTGALSTRLATDAAQVQATGTRL	831
QY	835	AVITQIANLGTGIIISLIYQWLTLLLAIVPIIAIAGVEMKMLSGQALKKKLEGA	894
Db	832	ALIAQNTANLGTGIIISLIYQWLTLLLSVVPFIAVSGIVEMKMLAGNAKRDKMEMEA	891
QY	895	GKIATEAIENFRVVSILTREQFEYMYAQSLOVPYRNSLRKAHIFGVFSITQAMMYFSY	954
Db	892	GKIATEAIENIRTVSILTQERKFESYVEKLHEPYRNSYRKAHYGITISQAFMYFSY	951
QY	955	ACGCFRGAYLVANEENFODVLVFSIAIVFGAMAVQVSSFAPDYAKAKVSAAHVIMIE	1014
Db	952	ACGCFRGAYLVINGHMRFRDVLVFSIAIVFGAVALGHASSFAPDYAKAKLSAAYLSLPE	1011
QY	1015	KSLPIDSYPHGLKPNLTLEGNVTNFVFNYPTRPDIPVLQGLSLEVKKGQIALVGS	1074
Db	1012	RQPLIDSYSREGMWPDKFEGSVTFNEVFNYPTRANPVLQGLSLEVKKGQIALVGS	1071
QY	1075	CGKSTVVQLLERYDPLAGSVLDGKEIKHLNVQWLRHLGIVSQRPILFDCSIAENIAY	1134
Db	1072	CGKSTVVQLLERYDPMAGTVLDDGQAKKLNVQWLRHLGIVSQRPILFDCSIAENIAY	1131

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: glycoprotein; membrane protein; nucleotide binding; P-loop
F:455-462/Region: nucleotide-binding motif A (P-loop)

Query Match 49.5%; Score 3201; DB 2; Length 1321; Best Local Similarity 50.2%; Pred. No. 3.5e-179; Matches 646; Conservative 235; Mismatches 370; Indels 36; Gaps 11;	
QY 20 KKSKEKKEKKP--TVSTFAMFRYSNWLDRMLYVGTMAAIIHGAALPLMLVFGNMTDS 77	
DB 30 KKSRLQDKMKEGDIRVGFELFRSSKDIWLMLMGVCALLHGMAGPQGLIIFGIMTDI 89	
QY 78 FANAGISR-----NKTFFVLINES----ITNNTQHFNLHEEMTYYYSGI 122	
DB 90 FIKYDIERQLEIPGKACVNTI--VWINSFHONMTNGTVCGLVDIESMIKFSGIYAGV 148	
QY 123 GAGVLVAAYIQVSWFCLAAAGRKQIRKQFFHAIMRQEGWFDVHDVGLNLTLDVDSK 182	
DB 149 GMTVLILGYFQIRLWITGARQIRRMKIFYFRIMRMEIGWFDCTSVGELNSRFADDIK 208	
QY 183 INEGIGDKVGMFPQSIAFTFTGFTGKLTVLILAIQVGLSAIAWAKILSSFTD 242	
DB 209 INDIAIDQLAHFLQRMSTAMCGLLGFIYRGWKLTLVLAVSPILGIGAAVIGLSIAKFFE 268	
QY 243 KELLAYAKAGAAVEEVLAAIRTVIAFGGQKKELERYKNLEAKRIGIKKAITANISGA 302	
DB 269 LELKAYAKAGSIADENVLSLRTVAAFGENKEVEREKNLVFAQRWGIKGMVWGFTGY 328	
QY 303 AFLLIYASALATFWYGTSLVL--SSEYTIQGLTVFFSVLIGAFSGOASPIEAFANARG 361	
DB 329 MWCLIFFCYALAFWYGSTLVLDDEEYTPGTLVQIFLCVILAAMNIGHASSCLEIFSTGCS 388	
QY 362 AAYEIFKIIDNKPISIDYSKGHKPNIDKNLEFKNVHESYPSRKEVKILKGLNLAVQSG 421	
DB 389 AATNIFQITIDROPYIDCMGSDGKLDRIKGEIEFHNFTHFYPRDPVKIILNLSMWIKPG 448	
QY 422 QTVALVNSGGCKSTTVOLMORLYDPTDGMVCIDGQDIRTVNRHLREITGVYSQBPVLF 481	
DB 449 ETTALVSSGNGAKSTALQILQRFYDPCGVMVLDGHDIRSLNIRLWDQIGIVEQBPVLF 508	
QY 482 ATTIAENIRYGRNWTDEIEKAVKENAYDFIMKLPNKFDTLVGERGARGLSGQKORIA 541	
DB 509 STTIAENIRFREDATMEDIVQAADKANAYINIMALPQOQDTLVGEGGQMSGGQKORVA 568	
QY 542 IARALVRNPKILLDEATSDLTSEAVVOALDKARKGRTTVIAHRLSTVRNADVIAG 601	
DB 569 IARALIRNPKILLDMATSLDNSEARVQEAALKIIGHITLSVAHRLSTVRADVIIG 628	
QY 602 FDDGVIVEKGNHDELMKEGYIKLVMTQTRNGEIELENA--TGESKSESDALEM-----S 655	
DB 629 FEHGVAVERGTHEELLERKGVYFVLVTILQSQGDNAHKETSMGKDATFEGGTLETFSRGS 688	
QY 656 PKDSGSSLIKRSTYR---SIHAP--OGODPKLTGED----LNENVPVSFWRIKLKN 705	
DB 689 YRDSLRASIRQSKQSLSLLPHDPLAVADHKSSYKOSKONDVLVEEVEAPVRILKYN 748	
QY 706 STEPFFVVGIFCAIINGLOPAFSIIFSRIGITFTRDEDEPETKRONSMFSLFLVLGI 765	
DB 749 IPEWHYILVGSLSAINGAVTPYISLLFSQLGTFSL-LDKEQQRSEIHSMLCFVLIGC 807	
QY 766 ISFTIFFLQGTFFKAGELIKRLRYWVFRSMLRQDQSWFDDPNKTTGALTTRLANDAAQ 825	
DB 808 VSIQTFQLQGYTFKSGELLTKRLKFGFKAMLGQDGFWDLDRNPGVLTTRLATDASQ 867	
QY 826 VKGATGSLAVTONIANLGTGIISLIYGWOLTLILLAIVIPITAIAGVEMKMLSGOAL 885	
DB 868 VOGATGSGVMKVNSTNIIAALLIAFFFSWKLSLIITIFFPFLALSGAVTKMLTGAS 927	
QY 886 KDKKELEGAKIATBAENFRVSLTRERQKPEYNYAQSGLQVPRYNSLRKAHIFGVSEFI 945	
DB 928 QDKQALEKAGQITSEALSNIRTVAGIGVEGRFIRAFEVELQTSYKTAVRKANIYGLCFAF 987	
QY 946 TQAMMYFSYAGCFRGAYLVANEFNFDQVLLVFSVAIVFGAMAGVQSSFPADYKAKVKS 1005	

DB 988 SGIATFLANSAAAYRGYGLIAYEGLGFSHFRVYSSVVLATAVGRFTSTPSPYAKAKIS 1047	
QY 1006 AAHVIMIEKSPLDISYSPHGLKPKNTLEGNVTEVFNYPTRDIPVLOGLSLEVKKGO 1065	
DB 1048 AARFQLDLDRKPPINVYSEAGEKWNFGKIDFIDCKFTYPSRDPDIQVNLGSLSVNPGQ 1107	
QY 1066 TLALVSGSGCKSTVWQVLLERFYDPLAGSVLIDGKEIKHLNVQWLRAHLGIVSQEPILFD 1125	
DB 1108 TLAFVSGSGCKSTSIQLERFYDQGTVMIDGSKVKNIQFLRSNIGIVSQEPVLFD 1167	
QY 1126 CSIAENIAYGNSRVVSHHEETMQAAKANIHHFETIETPEKYNTRVGDGKGTQLSGGQKORI 1185	
DB 1168 CSIMDNIKYGDNTKEISVERAIAAAKQALHDFVMSLPEKYETNVGIOGSOLESAGEKORI 1227	
QY 1186 AIARALVRQPHILLDEATSDLTSEKVVQEQALDKAREGTCIVIAHRLSTIQNADLV 1245	
DB 1228 AIARAIVRDPKILLDEATSDLTSEKVTQVTDLKAREGTCIVIAHRLSTIQNSDIIA 1287	
QY 1246 VFQNGKVKREHGTHTOOLLAQKGIYFSMI 1272	
DB 1288 VYSGGVVIEKGTHEKLMQAQKGIYKLV 1314	
RESULT 13	
T42842	
bile salt transport protein, ATP-dependent - rat	
N:Alternate names: bile salt export pump, sister of P-glycoprotein	
C:Species: Rattus norvegicus (Norway rat)	
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-May-2000	
C:Accession: T42842	
R:Gerloff, T.; Stieger, B.; Hagenbuch, B.; Madon, J.; Landmann, L.; Roth, J.; Hofmann	
J. Biol. Chem. 273, 10046-10050, 1998	
A:Title: The sister of P-glycoprotein represents the canalicular bile salt export pump	
A:Reference number: 222272; MUID:98212048	
A:Accession: T42842	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-1321 <GPR>	
A:Cross-references: EMBL:U69487; NID:g3075421; PID:g3075422; PIDN:AAC40084.1	
A:Experimental source: strain Prague-Dawley; liver	
C:Genetics:	
A:Gene: spgp	
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology	
C:Keywords: ATP; duplication; glycoprotein; membrane protein; P-loop	
Query Match 49.4%; Score 3197; DB 2; Length 1321; Best Local Similarity 50.1%; Pred. No. 5.9e-179; Matches 645; Conservative 235; Mismatches 371; Indels 36; Gaps 11;	

QY 20 KKSKEKKEKKP--TVSTFAMFRYSNWLDRMLYVGTMAAIIHGAALPLMLVFGNMTDS 77	
DB 30 KKSRLQDKMKEGDIRVGFELFRSSKDIWLMLMGVCALLHGMAGPQGLIIFGIMTDI 89	
QY 78 FANAGISR-----NKTFFVLINES----ITNNTQHFNLHEEMTYYYSGI 122	
DB 90 FIKYDIERQLEIPGKACVNTI--VWINSFHONMTNGTVCGLVDIESMIKFSGIYAGV 148	
QY 123 GAGVLVAAYIQVSWFCLAAAGRKQIRKQFFHAIMRQEGWFDVHDVGLNLTLDVDSK 182	
DB 149 GMTVLILGYFQIRLWITGARQIRRMKIFYFRIMRMEIGWFDCTSVGELNSRFADDIK 208	
QY 183 INEGIGDKVGMFPQSIAFTFTGFTGKLTVLILAIQVGLSAIAWAKILSSFTD 242	
DB 209 INDIAIDQLAHFLQRMSTAMCGLLGFIYRGWKLTLVLAVSPILGIGAAVIGLSIAKFFE 268	
QY 243 KELLAYAKAGAAVEEVLAAIRTVIAFGGQKKELERYKNLEAKRIGIKKAITANISGA 302	
DB 269 LELKAYAKAGSIADENVLSLRTVAAFGENKEVEREKNLVFAQRWGIKGMVWGFTGY 328	
QY 303 AFLLIYASALATFWYGTSLVL--SSEYTIQGLTVFFSVLIGAFSGOASPIEAFANARG 361	
DB 329 MWCLIFFCYALAFWYGSTLVLDDEEYTPGTLVQIFLCVILAAMNIGHASSCLEIFSTGCS 388	

Db 661 KPGAPPAPAAEKEIKRLKKELEEBGAVKANLFKILYARPENWIVYFFAIIAALIOGAVM 720
QY 727 PAFSIIIFSRIGIFTRDEDEPKYKRONSMFSLVLVLGIISIFITFFLOGFEGKAGEILT 786
Db 721 PAFSLFQSILINVS-NPDRDQMKDGHFWALMFLVLAAVQGTSMFQCSLFGVAAERLT 779
QY 787 KRLRYMVFSLRQDVSFDDPKNTGALTTLANDAAQVKAIGSRSLAVITQNTIANLGT 846
Db 780 MRLRSKYRNVLRQDQATYEDMPKHSGRITTRLATDAPNIKSAIDRYLGSIFNATASVGG 839
QY 847 GIISIIYQWQLTLLLAIVPIIAIAGVVMKMLSGQALKDKELEGAGKIATEAIENPR 906
Db 840 GLGIAFYQWQAFVMAIFPEMAYVQALMVKYHGSATSDAKEMENAGKTAMEAIENIR 899
QY 907 TVVSLTREOKFEWYAQSLOVYRNSLRKAHIFGVFSFISITQAMVYFAGCFEGAYLV- 965
Db 900 TVQALTLQTKLYNIFCSHLDAPHGGNISKAIIRGLTYGFANSIQFTFYAAAFRGLFLIF 959
QY 966 -ANEFMNFODVLVFAIVFGAMAVGVSVFAPDYAKAKVSAAHVIMITEKSLPLDISYSP 1024
Db 960 DKNVLMPEPNVLRVLAISFSGTIGFAASYPEYIKATFAAGLNFMLEEPRIDGMTS 1019
QY 1025 HGLKPNTEGNTVFNENVPYRPRDIPVLQGLSLEVKKGTQIALVSGSCGKSTVQQL 1084
Db 1020 SGTYPQ-LSGEVKNLVFFRYPERPAPVILQGLNVHVKPGQTIALVGPSCGKSTVISLL 1078
QY 1085 ERYDPLAGSLVDGKEIKHLNVOMLRAHLGIVSQPILPDCSIAENIAYGNSRVVSH 1144
Db 1079 ERLYDPLEGAVTVNDNLRQMPKHLKHALVQSPEILFDSIRENIVYGLQPEYTHE 1138
QY 1145 EIMQAKENIHFIEIETLPKYNTRVGDGTQGLSGGQKQRIATARALVPROPHILLDEAT 1204
Db 1139 QIETACSKANIHKFIDELPDYETRVGKGTQGLSGGQKQRIATARALIRNPKILLDEAT 1198
QY 1205 SALDTESEKVVQALDKARGRCIVIAHRLSTIQNALDILVFPQNKVKEHGHQOLLAQ 1264
Db 1199 SALDTESEKQVQALDAAKDRCTIVVAHRLSTIVNAGCIMVYGVQVGEQGHNELIAK 1258
QY 1265 KGIYFMSISVQA 1276
Db 1259 RGAYFALTQKQS 1270

RESULT 15
T31073
multidrug resistance p-glycoprotein - nematode (Haemonchus contortus)
C:Species: Haemonchus contortus
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T31073
R;Xu, M.; Molento, M.; Blackhall, W.; Ribeiro, P.; Beech, R.; Prichard, R.
Mol. Biochem. Parasitol. 91, 327-335, 1998
A:Title: Ivermectin resistance in nematodes may be caused by alteration of P-glycoprotein
A:Reference number: 220968; MUID:98226176
A:Accession: T31073
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1275 <XUM>
A:Cross-references: EMBL:AF003908; NID:G3057041; PID:G3057042; PIDN:AAC38987.1
A:Experimental source: developmental stage adult
C:Genetics:
A:Gene: PGP-A
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology

Query Match 43.5%; Score 2810; DB 2; Length 1275;
Best Local Similarity 44.1%; Pred. No. 2.6e-156;
Matches 565; Conservative 274; Mismatches 415; Indels 26; Gaps 12;

QY 8 KGSAEKNFWKKSSK-----EKKKKPTVSTFAMFRYSNWLDRMLVGTMAAIIHGA 62
Db 4 QGQDDERPLLGSSKSSIGEVSKKEPTTINRGILSLATLTDYVLLAGTAPCVHGA 63
QY 63 ALPLAMLVFGNNMTDFANAGISRNKTFPVIINESITNNQHFINHLEEMT-----YAY 118

Db 64 GFSVLGIVLGMVTVFLRA--QNSEF--VLGTVSRDPEGLPALTKKEEDTLVRRVCLY 117
QY 119 YSGIGAGVLAAYIQVSWFCLAAGRQILKIRKOFFHAIMRQIEGWDFVHDVGEINTRLTD 178
Db 118 YLGIGFAMFATSYIQVCMTEFAERITHKRLKYLKAILRQIQSWFIDQIQOTGNLTARLTD 177
QY 179 DVSKINGIGDKGVQMFTQSTATFTFTGIVGFTPGWKLTIVLILSAPVLGISAALIAKILS 238
Db 178 DLERVREGLDGKLSLFIQMVSAFVAGFCVGFAYSWSTLVMVVPFIVISANMWSKIVA 237
QY 239 SFTDKELLAYAKAGAAVEELAAITVIAFGQKKELERNKLNLEAKRIGIKKAITANI 298
Db 238 TRTQVEQETAVAGAIABETFSSTRTVHSICGKRELTREAALEKGRQGLVKYFVMGV 297
QY 299 SIGAAFLIIYASALAFWYGTSLVSS-ETVIGQVLVFFSVLIGAFISIQASPSIEAFA 357
Db 298 GVGFQMCYTVSYALAFWYGVLLINDPALDRGRIFTVFFFAVMSGSAALCTCLPHLNTIS 357
QY 358 NARGAAYEIFKLIIDNKSIDSYSGHKKPDNIKGNLFKNVHFSYPSRKEVKILKGNLKL 417
Db 358 IARGAVRSVLVINSRPKIDPYSLDGIVLNNMGRSIRFNKVNHFSPSRRLQILKGVSLQ 417
QY 418 VOSGOTVALVGNCGSGKSTTVQLMORLYDPTDGMVCIDGODIRTINVRHLREITGVVSQE 477
Db 418 VSAGOKIALVSGSCGCKSTNVNLLRFYDTPRGKVTDDIDVCDLVNOKLREQIGVVSQE 477
QY 478 PVLFATTIENIRYGRNVTMDEIEKAVKEANAYDFIMKLPNKEDTLVGERGABLSGOK 537
Db 478 PVLFDGTLFENIKMGYEQATMEEQEACRVANAADFTKRLPEGYGTRVGERGVSGGQK 537
QY 538 ORIATARALVNPKTLILDDEATSLDTESEAVOVVALDKARKGRTTIVIAHRLSTVRNAD 597
Db 538 ORIATARALVNPKTLILDDEATSLDTESEAVOVVALDKARKGRTTIVIAHRLSTVRNAD 597
QY 598 VIAGDDGVIVVEKGNHDELMKEGIYFKLVMTQTRNGEIELEENATGESKSSDALEMSPK 657
Db 598 QIVFVKNGTIVQGTTHAELMKNRGVFFEMTQAVLRQKEEVEVLDSDAESDVSPDIALP 657
QY 658 DSGSLLIKRRSTRRSIHA-POGQDRKLTGKEDLNENVSFSVFWRLTKLNSTWEPYFVVG 716
Db 658 HLSSLRKSTRSALSAVPSVRSQI-EMEDLRAPKTPMS--KIFYFNKDKWGFILGL 714
QY 717 FCALINGGLQPAFSIIFRIGIFTRDEDEPKYKRONSMF-SVLFLVLGIISITFFFLQ 775
Db 715 IACILITGVTPTFAVLYAQIIQVYS---EPVDMQGHVLFWCAGAFIVGLVHAFAPFSA 771
QY 776 FTFGKAGEILTRKRYMVFSLRQDVSFDDPKNTGALTTLANDAAQVKAIGSRLA 835
Db 772 ICLGRGGEALTCKLREAFKNLLRQNVGVFYDDIRHGTGKLCRTFRATDAPNVR-YVFTRLP 830
QY 836 VITQNIANLGTGIISLIVGWLTLILLAIPTIAGVVMKMLSGQALKDKELEGAG 895
Db 831 GVLSSVVTIIGALVIGTFPGWQALILVMVPLIIGSGYFEMRMQFGKKMRDTELEEGAG 890
QY 896 KIATEAIENFRTVSLTREQKFEYMYAQSLOVYRNSLRKAHIFGVFSITQAMVYFSA 955
Db 891 KVASQAVENIRTVHALNRQEQHFHYCEYLKEPYRENLCQAHTYGVGFYAFQSLLFFMYA 950
QY 956 GCFRFGAYLVANEFNFQDVLVFAIVFGAMAVGVSVFAPDYAKAKVSAAHVIMITEK 1015
Db 951 VAFWIGAIIVDNHSMQPIDVYRVFFAFMFCGQMVGNISFIFDVVKARLAASLLYLIH 1010
QY 1016 SPLDISYSPHGLKPNTEGNTVFNENVPYRPRDIPVLQGLSLEVKKGTQIALVSGSCG 1075
Db 1011 PSEIDNLSEGV-TKKISGHISFRNVYFNPTRRQIRVLRGLNLEINPGTTVVALVQSGC 1069
QY 1076 GKSTVVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRHAHLGIVSQPILPDCSIAENIAYG 1135
Db 1070 GKSTVALLERFYNQNKGVITVDGENIRNMRNLREQVICVSOEPTLEDCTIMENICYG 1129
QY 1136 DNSRVVSHBEIMOAAKEANIHFIETLPKYNTRVGDGTQGLSGGQKQRIATARALVROP 1195
Db 1130 LDDPKFSYEQVAAAKMANIHNFVLGPEGYDTRVGEKGTQGLSGGQKQRIATARALIRP 1189

QY 1196 HILLDEATSALDTESEKVOEALDKAREGRTCIIVIAHRLSTIONADLIVVFONGKYKEH 1255
Db 1190 PILLDEATSALDTESEKIVQDALEVARQGRTCIVIAHRLSTIQSDVIVMIQEGKATDR 1249
QY 1256 GTHQOLLAQKGIYFSMISVQ 1275
Db 1250 GTHEHLLMKNDLYKRLCETQ 1269

Search completed: November 6, 2002, 18:44:37
Job time : 28.0473 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:30:44 ; Search time 10.4899 seconds
(without alignments)
4724.646 Million cell updates/sec

Title: US-09-672-725C-4

Perfect score: 6467

Sequence: 1 MDPGGRKGSAEKNFWMGK.....LLAQGIYFMSISVQAGAKR 1280

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5859.5	90.6	1280	MDR1_HUMAN	P08183 homo sapien
2	5661	87.5	1276	MDR1_CRIGR	P21448 cricetus
3	5631.5	87.1	1276	MDR3_MOUSE	P21447 mus musculus
4	5304.5	82.0	1276	MDR1_MOUSE	P06795 mus musculus
5	5281.5	81.7	1276	MDR2_CRIGR	P21449 cricetus
6	5236	81.0	1277	MDR1_RAT	P43245 rattus norv
7	4898	75.7	1279	MDR3_HUMAN	P21439 homo sapien
8	4850.5	75.0	1276	MDR2_MOUSE	P21440 mus musculus
9	4810	74.4	1281	MDR3_CRIGR	P23174 cricetus
10	4796.5	74.2	1278	MDR2_RAT	P08201 rattus norv
11	3292	50.9	1321	AB11_HUMAN	O95342 homo sapien
12	3219	49.8	1321	AB11_RABIT	O9n0v3 oryctolagus
13	3201	49.5	1321	AB11_MOUSE	O9n0v3 mus musculus
14	3197	49.4	1321	AB11_RAT	O70127 rattus norv
15	2738	42.3	1321	MDR1_CAEL	P34712 caenorhabdi
16	2561.5	38.6	1302	MDR4_DROME	Q00449 drosophila
17	2491	38.5	1302	MDR5_DROME	Q00748 drosophila
18	2347.5	36.3	1254	MDR3_CAEL	P34713 caenorhabdi
19	2267.5	35.1	1362	MDR1_SCHPO	P36619 schizosacch
20	2071	32.0	1280	MDR1_LEIEN	Q06034 leishmania
21	1549.5	24.0	1419	MDR_PLAFF	P13568 plasmodium
22	1223.5	18.9	1290	ST6E_YEAST	P12866 saccharomyc
23	1128.5	17.5	1336	MAM1_SCHPO	P78966 schizosacch
24	1026.5	15.9	1323	HST6_CANAL	P53706 candida alb
25	887	13.7	735	ABC8_HUMAN	O9nut2 homo sapien
26	827	12.8	582	MSBA_ECOLI	P27299 escherichia
27	819	12.7	587	MSBA_HAEIN	P44407 haemophilus
28	770	11.9	598	Y288_THENA	Q9wyc4 thermotoga
29	769.5	11.9	820	MDL2_YEAST	P33311 saccharomyc
30	769	11.9	1437	MRP5_HUMAN	O15440 homo sapien
31	764.5	11.8	1436	MDP5_RAT	O9qym0 rattus norv
32	763	11.8	695	MDL1_YEAST	P33310 saccharomyc
33	754.5	11.7	1436	MRP5_MOUSE	Q9rix5 mus musculus

34 746 11.5 1545 1 MRP2_HUMAN Q92887 homo sapien
35 742.5 11.5 726 1 YFX9_SCHPO Q997m7 schizosacch
36 730.5 11.3 1564 1 MRP2_RABIT Q28689 oryctolagus
37 728.5 11.3 1325 1 MRP4_HUMAN O15439 homo sapien
38 723.5 11.2 703 1 TAP2_RAT P36372 rattus norv
39 721.5 11.2 685 1 MDL1_CANAL P97998 candida alb
40 716 11.1 1541 1 MRP2_RAT O63120 rattus norv
41 709.5 11.0 607 1 HEPA_ANASP P22638 anabaena sp
42 709 11.0 575 1 YWJA_BACSU P45861 bacillus su
43 707 10.9 584 1 LMRA_LACLA Q9chl8 lactococcus
44 700.5 10.8 604 1 YFIC_BACSU P54719 bacillus su
45 695 10.7 584 1 LMRA_LACLC P97046 lactococcus

ALIGNMENTS

RESULT 1
MDR1_HUMAN
ID MDR1_HUMAN STANDARD; PRT: 1280 AA.
AC P08183; Q12755; Q14812;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein 1).
GN ABCB1 OR PGY1 OR MDR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87028230; PubMed=2876781;
RA Chen C.-J., Chai J.E., Ueda K., Clark D.P., Pastan I., Gottesman M.M.,
RA Roninson I.B.;
RT "Internal duplication and homology with bacterial transport proteins
RT in the mdr1 (P-glycoprotein) gene from multidrug-resistant human
RT cells.";
RL Cell 47:381-389(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90094448; PubMed=1967175;
RA Chen C.-J., Clark D.P., Ueda K., Pastan I., Gottesman M.M.,
RA Roninson I.B.;
RT "Genomic organization of the human multidrug resistance (MDR1) gene
RT and origin of P-glycoproteins.";
RL J. Biol. Chem. 265:506-514(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97190336; PubMed=9038218;
RA Chen G., Duran G.E., Steger K.A., Lacayo N.J., Jaffrezou J.P.,
RA Dunonnet C., Sicic B.I.;
RT "Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein,
RT altered phenotype, and resistance to cyclosporins.";
RL J. Biol. Chem. 272:5974-5982(1997).
RN [4]
RP SEQUENCE OF 1-234 FROM N.A.
RA Smith A., Beck C., Gibson A.;
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 178-215 AND 800-856 FROM N.A.
RX MEDLINE=90290529; PubMed=1972623;
RA Gekeler V., Weger S., Probst H.;
RT "mdr1/P-glycoprotein gene segments analyzed from various human
RT leukemic cell lines exhibiting different multidrug resistance
RT profiles.";
RL Biochem. Biophys. Res. Commun. 169:796-802(1990).
RN [6]
RP SEQUENCE OF 1-23 FROM N.A.
RA Kioka N., Tsubota J., Kakehi Y., Komano T., Gottesman M.M.,
RA Pastan I., Uedak K.;
RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED

CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC -|- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, SMALL INTESTINE
CC AND BRAIN.
CC
CC -|- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14758; AAA59575.1; -;
CC EMBL; M29447; AAA59576.1; -;
CC EMBL; M29424; AAA59576.1; JOINED.
CC EMBL; M29425; AAA59576.1; JOINED.
CC EMBL; M29426; AAA59576.1; JOINED.
CC EMBL; M29427; AAA59576.1; JOINED.
CC EMBL; M29428; AAA59576.1; JOINED.
CC EMBL; M29429; AAA59576.1; JOINED.
CC EMBL; M29430; AAA59576.1; JOINED.
CC EMBL; M29431; AAA59576.1; JOINED.
CC EMBL; M29432; AAA59576.1; JOINED.
CC EMBL; M29433; AAA59576.1; JOINED.
CC EMBL; M29434; AAA59576.1; JOINED.
CC EMBL; M29435; AAA59576.1; JOINED.
CC EMBL; M29436; AAA59576.1; JOINED.
CC EMBL; M29437; AAA59576.1; JOINED.
CC EMBL; M29438; AAA59576.1; JOINED.
CC EMBL; M29439; AAA59576.1; JOINED.
CC EMBL; M29440; AAA59576.1; JOINED.
CC EMBL; M29441; AAA59576.1; JOINED.
CC EMBL; M29442; AAA59576.1; JOINED.
CC EMBL; M29443; AAA59576.1; JOINED.
CC EMBL; M29444; AAA59576.1; JOINED.
CC EMBL; M29445; AAA59576.1; JOINED.
CC EMBL; M29446; AAA59576.1; JOINED.
CC EMBL; AF016535; AAB69423.1; -;
CC EMBL; AC002457; AAC82531.1; -;
CC EMBL; M37724; AAA88047.1; -;
CC EMBL; M37725; AAA88048.1; -;
CC EMBL; X58723; CAA41558.1; -;
CC PIR; A25059; DVHU1
CC PIR; A34914; A34914.
CC MIM; I71050; -;
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001140; ABC_transporter_tmemb.
CC InterPro; IPR003439; ABC_transporter.
CC InterPro; IPR001687; ATP_GTP_A.
CC Pfam; PF00664; ABC_membrane; 2.
CC Pfam; PF00005; ABC_tran; 2.
CC SMART; SM00382; AAA; 2.
CC PROSITE; PS00211; ABC_TRANSPORTER; 2.
CC ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
CC Multigene family.
CC DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
CC TRANSSEM 52 72 POTENTIAL.
CC TRANSSEM 120 140 POTENTIAL.
CC TRANSSEM 189 209 POTENTIAL.
CC TRANSSEM 216 236 POTENTIAL.
CC TRANSSEM 297 317 POTENTIAL.
CC TRANSSEM 326 346 POTENTIAL.
CC DOMAIN 347 710 CYTOPLASMIC (POTENTIAL).
CC TRANSSEM 711 731 POTENTIAL.
CC TRANSSEM 757 777 POTENTIAL.
CC TRANSSEM 833 853 POTENTIAL.
CC TRANSSEM 854 874 POTENTIAL.
CC TRANSSEM 937 957 POTENTIAL.
CC TRANSSEM 974 994 POTENTIAL.
CC -DOMAIN 995 1280 CYTOPLASMIC (POTENTIAL).
CC NP_BIND 427 434 ATP (BY SIMILARITY).

FT	NP_BIND	1070	1077	ATP (BY SIMILARITY).
FT	REPEAT	1	637	
FT	REPEAT	638	1280	
FT	CARBOHYD	91	91	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	94	94	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	99	99	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	23	23	S -> R (IN REF. 6).
FT	CONFLICT	185	185	G -> V (IN REF. 1 AND 3).
FT	CONFLICT	336	336	MISSING (IN REF. 3).
FT	CONFLICT	412	412	G -> A (IN REF. 3).
FT	CONFLICT	438	438	Q -> S (IN REF. 3).
FT	SEQUENCE	1280 AA;	141462 MW;	ABIC279531F43675 CRC64;
Query Match				
Best Local Similarity	90.6%;	Score 5859.5;	DB 1;	Length 1280;
Matches 1159;	Conservative	Pred. No. 0;		
			Mismatches 60;	Indels 3; Gaps 3;
Qy	1	MDPGRKGSA-EKNFMKMGKSKKEKKPTVSTFAMERYSNWLDRLMLVGTMAAII	59	
Db	1	MDLEGDRNGGAKKNFFKLNNKSEKKKPTVSFMRISNWLDRLLVWVGTLLAAII	60	
Qy	60	HGAALPLMLVFGNMTDSFANAGISRNKTFVFIINESITNTNQHFINHLEEMTYAYYY	119	
Db	61	HGAGLPLMLVFGMTDIFANAG-NLEDLMSNITNRSNDINDTGFFMN-LEEDMTRYAYYY	118	
Qy	120	SGIGAGVLVAAYIQVSFWCLAGROILKIRKQFFHAIMRQBIGWFDVHDVDELNRLTDD	179	
Db	119	SGIGAGVLVAAYIQVSFWCLAGROILKIRKQFFHAIMRQBIGWFDVHDVDELNRLTDD	178	
Qy	180	VSKINEGIDKVGMPFQSIATFTFGIVGTPGWKLTIVILAIISPVGLSLAAIWAKILSS	239	
Db	179	VSKINEGIDKVGMPFQSMATFTFGIVGTPGWKLTIVILAIISPVGLSLAAIWAKILSS	238	
Qy	240	FTDKELLAYAKAGAAVEEVLAAIRTVIAFGQKKELERYNKNLEAKRIGIKKAITANIS	299	
Db	239	FTDKELLAYAKAGAAVEEVLAAIRTVIAFGQKKELERYNKNLEAKRIGIKKAITANIS	298	
Qy	300	IGAAFLLIYASALAFWYGTSLVLSSEYTIQGLTVFVSVLIGAFSGQASPSFEAFANA	359	
Db	299	IGAAFLLIYASALAFWYGTSLVLSSEYTIQGLTVFVSVLIGAFSGQASPSFEAFANA	358	
Qy	360	RGAAVEIEFKIDNKPSIDSYSGHGPDKINIKGNLFKFNHFSYPSRKEVKILKGLNKVQ	419	
Db	359	RGAAVEIEFKIDNKPSIDSYSGHGPDKINIKGNLFKFNHFSYPSRKEVKILKGLNKVQ	418	
Qy	420	SGQTVALVNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDIRTINVRHLREITGVVQSEPV	479	
Db	419	SGQTVALVNSGCGKSTTVQLMQRLYDPTDGMVCDIGQDIRTINVRHLREITGVVQSEPV	478	
Qy	480	LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGARGSGQKOR	539	
Db	479	LFATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGARGSGQKOR	538	
Qy	540	IAIARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI	599	
Db	539	IAIARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVI	598	
Qy	600	AGFDGVIIVEKGNHDELMKEGIYFKLVTMOTRGNEIELEENATGSKESDALEMSPKDS	659	
Db	599	AGFDGVIIVEKGNHDELMKEGIYFKLVTMOTRGNEIELEENATGSKESDALEMSPKDS	658	
Qy	660	GSSLIKRRSTRRSIHAPOGQDRKLTGKEDLNENVPVSFWRLIKLNSTWEPYFVVGIFCA	719	
Db	659	RSSLIKRRSTRRSIHAPOGQDRKLTGKEDLNENVPVSFWRLIKLNSTWEPYFVVGIFCA	718	
Qy	720	IINGGLQPAFISFIISRIIGIFTRDEDPETKRONSNMFSVLFLVLGIIISFIIFLQGFEG	779	
Db	719	IINGGLQPAFISFIISRIIGIFTRDEDPETKRONSNLFSLLFLALGIIISFIIFLQGFEG	778	
Qy	780	KAGEILTTLRLYVNFPSMLRQDVSFDDPDKNTTGALTTRLANDAAQVKGAGISRLAVITQ	839	
Db	779	KAGEILTTLRLYVNFPSMLRQDVSFDDPDKNTTGALTTRLANDAAQVKGAGISRLAVITQ	838	


```

:|||||
296 MGAFFLIITASYALAFWGTSLVISKEYSIGQVLFVFFAVLIGASIGQASPNIEAFANA 355
QY 360 RGAAYEIFKIIDNKPSIDSYKSGHRPNKIKGNLEPKNVHFSYPSRKEVKILKGLNKVQ 419
Db 356 RGAAYEIFNIIDNKPSIDSFSGKNGYKPNKIKGNLEPKNIHFHSYPSRKQVILKGLNKVQ 415
QY 420 SGQTVALVGNSSGCGKSTTVQLMORLDPTDGMVCIIDGQDIRTINVRHLREITGVVSOEVP 479
Db 416 SGQTVALVGNSSGCGKSTTVQLMORLDPTDGMVCIIDGQDIRTINVRHLREITGVVSOEVP 475
QY 480 LFATTIAENIRYGRNVTWDEIEKAVKEANAYDFIMKLPNKFTDLVGERGARGSKQKOR 539
Db 476 LFATTIAENIRYGRNVTWDEIEKAVKEANAYDFIMKLPNKFTDLVGERGARGSKQKOR 535
QY 540 IAIARALVRNPKILLDEATFSAIDTSEAVVQVALDKARKGRFTTIVIAHRLSTVRNADVI 599
Db 536 IAIARALVRNPKILLDEATFSAIDTSEAVVQVALDKARKGRFTTIVIAHRLSTVRNADII 595
QY 600 AGFDGVIYKGNHDELMKEKGIYFKLVTMOTRGNETELEENATGESKSDALEMSPKDS 659
Db 596 AGFDGVIYKGNHDELMKEKGIYFKLVTMOTAGNETELEENATGESKSDALEMSSKDS 655
QY 660 GSSLIKRRTSRSTHAPQGDQKGLTKEDLNENVPVSFWRILKLNSTEMPYFVVGIFCA 719
Db 656 ASSLIKRRTSRSTHAPQGDQKGLTKEDLNENVPVSFWRILKLNSTEMPYFVVGIFCA 715
QY 720 IINGLOPAISIFSRIGIIFTRDEDPETKRONSFVFLVILGIIISFTTFPLQGTFC 779
Db 716 IINGLOPAISIFSRIGIIFTRDEDPETKRONSFVFLVILGIIISFTTFPLQGTFC 775
QY 780 KAGELTKRLRYMYFRSMLRQDSWFDPKNTTGALTTRLANDAAQVKAIGSRLAVITQ 839
Db 776 KAGELTKRLRYMYFRSMLRQDSWFDPKNTTGALTTRLANDAAQVKAIGSRLAVITQ 835
QY 840 NIANLGTGIIISLYGWLTLALLAIPIIAGVEMKMLSGQALKDKKELSGAKKIAT 899
Db 836 NIANLGTGIIISLYGWLTLALLAIPIIAGVEMKMLSGQALKDKKELSGAKKIAT 895
QY 900 EAIENFTVSLTREQFEYMYAQSLQVPYRNSLRKAHIFGVFSITQAMMYFSYAGCFR 959
Db 896 EAIENFTVSLTREQFEYMYAQSLQVPYRNSLRKAHIFGVFSITQAMMYFSYAGCFR 955
QY 960 FGAYIVANEFNFDVLLVFSIAIVFGAMAVGVSSFPADYAKAKVSAHVIMIEKSPLI 1019
Db 956 FGAYIVANEFNFDVLLVFSIAIVFGAMAVGVSSFPADYAKAKVSAHVIMIEKSPSI 1015
QY 1020 DSYSPHGLKPNLTLEGNVTFNEVFNYPTRPDIPIVQLGLSLEVKKGOTLALVSGSGCKST 1079
Db 1016 DSYSPHGLKPNLTLEGNVTFNEVFNYPTRPDIPIVQLGLSLEVKKGOTLALVSGSGCKST 1075
QY 1080 VVOLLERYDPLAGSLVDGKEIKHLNVQWLRHLGIVSOEPILFDCSIAENIAYGDSNR 1139
Db 1076 VVOLLERYDPLAGSLVDGKEIKHLNVQWLRHLGIVSOEPILFDCSIAENIAYGDSNR 1135
QY 1140 VVSHEEIMQAAKEANIHFIEPLPEKYNTRVDRKGTQLSGQKQRIARALVRQPHILL 1199
Db 1136 VVSHEEIMQAAKEANIHFIEPLPEKYNTRVDRKGTQLSGQKQRIARALVRQPHILL 1195
QY 1200 LDEATFSAIDTSEAVVQVALDKARKGRFTTIVIAHRLSTIONADILIVFONGKVKHEGTHQ 1259
Db 1196 LDEATFSAIDTSEAVVQVALDKARKGRFTTIVIAHRLSTIONADILIVFONGKVKHEGTHQ 1255
QY 1260 QLLAQKGIYFMSISVQAGAKR 1280
Db 1256 QLLAQKGIYFMSISVQAGAKR 1276

RESULT 3
MDR3_MOUSE STANDARD: PRT: 1276 AA.
AC P21447:
DT 01-MAY-1991 (Rel. 18, Created)
```

```
FT TRANSMEM 708 728 POTENTIAL.
FT TRANSMEM 753 773 POTENTIAL.
FT TRANSMEM 829 848 POTENTIAL.
FT TRANSMEM 853 880 POTENTIAL.
FT TRANSMEM 942 961 POTENTIAL.
FT TRANSMEM 964 984 POTENTIAL.
FT DOMAIN 985 1276 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 423 430 ATP (BY SIMILARITY).
FT NP_BIND 1066 1073 ATP (BY SIMILARITY).
FT REPEAT 1 632
FT REPEAT 633 1276
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 526 527 QL -> HV (IN REF. 2 AND 3).
FT CONFLICT 939 939 F -> S (IN REF. 2 AND 3).
FT CONFLICT 1036 1036 F -> V (IN REF. 2 AND 3).
SQ SEQUENCE 1276 AA; 140754 MW; 75C71F33E1F58481 CRC64;

Query Match
Best Local Similarity 86.7%; Pred. No. 3,4e-311;
Matches 1111; Conservative 78; Mismatches 85; Indels 7; Gaps 3;

QY 1 MDEGGKSGSAEKKFWKMGKSKKEKKETPTVSTFAMFYSNWLRLYMLVGTMAIIH 60
Db 1 MELEEDLKGRAADKRFKMGKSKKEKKETPAVSILTFMFRYAGWLDRLYMLVGTMAIIH 60
QY 61 GAALPLMLLVGNMTDSFANAG-ISRNNKTPPVIINESITNTQHFNIHLEEMTTVAYY 119
Db 61 GVALPLMLLVGNMTDSFASGVNYSKST-----NSEADKRAMFAK-LEEEMTTVAYY 114
QY 120 SGIGAGVLVAYIQVSWFCLAAAGRLKIRKQFFHAIMRQIEGWDFVDHVGELNRLTDD 179
Db 115 TGIGAGVLVAYIQVSWFCLAAAGRLKIRKQFFHAIMNQEIGWDFVDHVGELNRLTDD 174
QY 180 VSKINEGIGDKVGNFFOSIATFTGTGTGTPGKTLVLTLAISPVLGSLAAIWKILSS 239
Db 175 VSKINEGIGDKVGNFFOSIATFTGTGTGTPGKTLVLTLAISPVLGSLAAIWKILSS 234
QY 240 FTDKELLAYAKAGAVEVLAARTVIAFGQKKEKELRYNKNLEEAIRIGIKKAITANTIS 299
Db 235 FTDKELHAYAKAGAVEVLAARTVIAFGQKKEKELRYNKNLEEAIRIGIKKAITANTIS 294
QY 300 IGAFLLIYASYALAFWYGTSLVLSSEYTTGQVLTVFFSVLIGAFSIGQASPSIEAFANA 359
Db 295 MGAFLLIYASYALAFWYGTSLVLSSEYTTGQVLTVFFSVLIGAFSIGQASPSIEAFANA 354
QY 360 RGAAYEFKIIDNKPSIDTSKSGHKPDNFKGNLEPKNFHFSYPSRKEVKILKGLNKVQ 419
Db 355 RGAAYEFKIIDNKPSIDTSKSGHKPDNFKGNLEPKNFHFSYPSRKEVKILKGLNKVQ 414
QY 420 SGQTVALVGNSSGCKSTTVQLMQRLYDPTDGMVCDIGQDQIRTNVRLHREITGVVSGEPV 479
Db 415 SGQTVALVGNSSGCKSTTVQLMQRLYDPTDGMVCDIGQDQIRTNVRLHREITGVVSGEPV 474
QY 480 LFAATTAEINRYGRENWYDEIEKAYKEANAYDFIMKLPNKFDTLYGERGARGSGQKOR 539
Db 475 LFAATTAEINRYGRENWYDEIEKAYKEANAYDFIMKLPNKFDTLYGERGARGSGQKOR 534
QY 540 IATARALVRNPKILLDEATSAIDTSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADVI 599
Db 535 IATARALVRNPKILLDEATSAIDTSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADVI 594
QY 600 AGPDDGVIVKGNHDELMKEKGYFKVLVTWQTRGNIELENATGSKSESDALEMSPKDS 659
Db 595 AGPDDGVIVKGNHDELMKEKGYFKVLVTWQTRGNIELENATGSKSESDALEMSPKDS 654
QY 660 GSSLIKRSTRSIIHAPQODKRLGKTKEDLNENPVSFWIRILKLNSTWEPYVVGIFCA 719
Db 655 GSSLIKRSTRSIIHAPQODKRLGKTKEDLNENPVSFWIRILKLNSTWEPYVVGIFCA 714
QY 720 IINGGLQPAFISIFSRIGIETDEDEPETKRONSNWFSVLFLVLGIIISFITFLQGFTEG 779
Db 715 IINGGLQPAFISIFSRIGIETDEDEPETKRONSNWFSVLFLVLGIIISFITFLQGFTEG 774
QY 780 KAGEIILTKRLRYMVFVSMRLRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSRSLAVITQ 839
Db 775 KAGEIILTKRLRYMVFVSMRLRQDVSFDDPKNTTGALTTRLANDAAQVKAIGSRSLAVITQ 834
QY 840 NIANLGTGIIISLIYQWLTLLALLAIVPIIATAGVVMKMLSGQALKDKKELLEGACKIAT 899
Db 835 NIANLGTGIIISLIYQWLTLLALLAIVPIIATAGVVMKMLSGQALKDKKELLEGACKIAT 894
QY 900 EATENFTVVSLETRQKFEYMYAQSLQVYPYRNSLRKAHFVGSFSTQAMTFFSVAGCFR 959
Db 895 EATENFTVVSLETRQKFEYMYAQSLQVYPYRNSLRKAHFVGSFSTQAMTFFSVAGCFR 954
QY 960 FGAYIVANFEMFNPQDVLVLSAIVFGAMAVQVSSFADPYAKAKVSAAHVIMIEKSPLI 1019
Db 955 FGAYIVANFEMFNPQDVLVLSAIVFGAMAVQVSSFADPYAKAKVSAAHVIMIEKSPLI 1014
QY 1020 DYSVSPHGLKPNLTLEGNWTFNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSSGCGKST 1079
Db 1015 DYSVSTOGLKPNLMEGNWTFNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSSGCGKST 1074
QY 1080 VOLLERFYDPLAGSVLIDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSTAEINAYGDNRS 1139
Db 1075 VOLLERFYDPMAGSVFLDGEIKELQVQWMLRAHLGIVSQEPILFDCSTAEINAYGDNRS 1134
QY 1140 VVSHEEIMAAKEANIHFIETLPEKYNTRVGDGKTQSLGGQKQRIATARALVRPHILL 1199
Db 1135 VVSHEEIMAAKEANIHFIETLPEKYNTRVGDGKTQSLGGQKQRIATARALVRPHILL 1194
QY 1200 LDEATSAIDTESKVVQVEALDKAREGTCIVIAHRLSTIONADLIWVQNGVKKEHGHQ 1259
Db 1195 LDEATSAIDTESKVVQVEALDKAREGTCIVIAHRLSTIONADLIWVQNGVKKEHGHQ 1254
QY 1260 QLLAOKGIYFSMISVOAGAKR 1280
Db 1255 QLLAOKGIYFSMISVOAGAKR 1275

RESULT 4
MDRL_MOUSE
ID MDRL_MOUSE STANDARD; PRT; 1276 AA.
AC P06795;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein 1).
GN ABCB1 OR PGL1 OR PGL1-1 OR MDR1 OR MDR1B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87028229; PubMed=3768958;
RA Gros P., Croop J., Housman D.;
RT "Mammalian multidrug resistance gene: complete cDNA sequence
indicates strong homology to bacterial transport proteins.";
RL Cell 47:371-380(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89367274; PubMed=2570420;
RA Raymond M., Gros P.;
RT "Mammalian multidrug-resistance gene: correlation of exon
organization with structural domains and duplication of an ancestral
gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:6488-6492(1989).
RN [3]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=91042535; PubMed=2248681;
RA Raymond M., Gros P.;
RT "Cell-specific activity of cis-acting regulatory elements in the
promoter of the mouse multidrug resistance gene mdr1.";
RN [4]
```


Db 1257 QLLAOKGIYFSM--VOAGAKR 1275

RESULT 5

MDR2_CRIGR

ID MDR2_CRIGR STANDARD; PRT; 1276 AA.

AC P21449;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Multidrug resistance protein 2 (p-glycoprotein 2).

GN PGV2 OR PGP2

OS Cricetus griseus (Chinese hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetidae; Cricetulus.

OX NCBI_TaxID=10029;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92135896; PubMed=1685679;

RA Endicott J.A., Sarangi F., Ling V.;

RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein gene family.";

RL DNA Seq. 2:89-101(1991).

RN [2]

RP SEQUENCE OF 622-1276 FROM N.A.

RX MEDLINE=88122132; PubMed=2893255;

RA Endicott J.A., Juranka P.F., Sarangi F., Gerlach J.H., Deuchars K.L., Ling V.;

RT "Simultaneous expression of two P-glycoprotein genes in drug-sensitive Chinese hamster ovary cells.";

RL Mol. Cell. Biol. 7:4075-4081(1987).

CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY CANNOT.

CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M60041; AAA68884.1; -;

DR EMBL; M17896; AAA37007.1; -;

DR PIR; B27126; DVHYZC.

DR InterPro; IPR003593; AAA.

DR InterPro; IPR001140; ABC_transporter_tmem.

DR InterPro; IPR003439; ABC_transportr.

DR InterPro; IPR001687; ATP_Grp_A.

DR Pfam; PF00664; ABC_membrane; 2.

DR Pfam; PF00005; ABC_tran; 2.

DR SMART; SM00382; AAA; 2.

DR PROSITE; PS00211; ABC_TRANSPORTER; 2.

KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;

KW Multigene family.

FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 52 72 POTENTIAL.

FT TRANSMEM 119 139 POTENTIAL.

FT TRANSMEM 188 208 POTENTIAL.

FT TRANSMEM 215 235 POTENTIAL.

FT TRANSMEM 296 316 POTENTIAL.

FT TRANSMEM 325 345 POTENTIAL.

FT DOMAIN 346 708 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 709 729 POTENTIAL.

FT TRANSMEM 755 775 POTENTIAL.

FT TRANSMEM 831 851 POTENTIAL.

FT TRANSMEM 852 872 POTENTIAL.

FT TRANSMEM 935 955 POTENTIAL.

FT TRANSMEM 972 992 POTENTIAL.

FT DOMAIN 993 1276 CYTOPLASMIC (POTENTIAL).

FT NP_BIND 426 433 ATP (POTENTIAL).

FT NP_BIND 1068 1075 ATP (POTENTIAL).

FT REPEAT 1 635

FT REPEAT 636 1276

SQ SEQUENCE 1276 AA; 141057 MW; 5096B1385628812D CRC64;

Query Match 81.7%; Score 5281.5; DB 1; Length 1276;

Best Local Similarity 80.1%; Pred. No. 2.2e-291;

Matches 1026; Conservative 129; Mismatches 119; Indels 7; Gaps 4;

QY 1 MDPEGRKGSKEKFWKMGKSKKEKKK-PTVSTFAMFRYSNWLDRMLVMTMAAII 59

DB 1 MEFEEDFSARADKDFLKMGRKSKKEKENPNVIGFMFYADMLDKLMLVLTAAVL 60

QY 60 HGAALPLMLVFGNMTDSFANAGISRNKTPVLIINSEITNNTQHFINHLEEMTTAYYY 119

DB 61 HGTSIPLMLVFGNMTDSFTKAETS---IWPNTNQEINNTTEVSGSLEEDMATYAYY 117

QY 120 SGIGAGVLVAAYIQVSWFCLAAAGRIKQKQFFHAIMRQIEIGWFDVHDVDELNRLTDD 179

DB 118 TGIGAGVLIVAYIQVSWFCLAAAGRIKQKQFFHAIMRQIEIGWFDVHDVDELNRLTDD 177

QY 180 VSKINEGIGDKVGMFFOSIATFTFGTGVTPGWKLTVLILASPVLGSAIAWAKILSS 239

DB 178 VSKINDGIGDKIGMFFOSIATFLAAFTVIGFISGWKLTVLILASPVLGSAIAWAKILSS 237

QY 240 FTDKELLAYAKAGAVEELAAITVIAFGOKKELERYKNLEAKRIGIKKAITANIS 299

DB 238 FTNKLQAYAKAGAVEELAAITVIAFGQNKELERYKNLEAKNVIKKAIVANIS 297

QY 300 IGAFLLIYASALAFWYGTSLVLSSEYTGQVLTFFSVLIGAFSGASPSIEAFANA 359

DB 298 IGIALVLLVYASALAFWYGTSLVLSSEYSGQVLTFFSVLIGAFSGHIANIEVFANA 357

QY 360 RGAAYELFKIIDNKPISDSYSGHGPDKNKGLEPKNVHFSYPSRKEVKILGLNLKVQ 419

DB 358 RGAAYELFKIIDNEPISDSYSGHGPDKNKGLEPKNVHFSYPSRSGIKILGLNLKVQ 417

QY 420 SGQTVLVGNSGCGKSTTVQLMORLYDPTDGMVCDIGQDIRTINVRHLREITGVVSQEPV 479

DB 418 SGQTVLVGNSGCGKSTTVQLMORLYDPTDGMVCDIGQDIRTINVRHLREITGVVSQEPV 477

QY 480 LPTATTAEINRYGRENVTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGALSGGQKOR 539

DB 478 LPTATTAEINRYGRENVTDEIEKAVKEANAYDFIMKLPNKFDTLVGERGALSGGQKOR 537

QY 540 IATARALVRNPKILLDEATSDALDTESEAVVOALDKARKGTTIVIAHRLSTVRNADVI 599

DB 538 IATARALVRNPKILLDEATSDALDTESEAVVOALDKARKGTTIVIAHRLSTVRNADVI 597

QY 600 AGFDGIVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGSKSESDALEMSPKDS 659

DB 598 AGFDGIVIVEKGNHDELMKEGIYFKLVMTQTRGNEIELENATGSKSESDALEMSPKDS 657

QY 660 GSSLIKRRTRSIHAPQGDRLKGTEDLNENSVSFWRIILKLNSTWEPYFVVGIFCA 719

DB 658 KPSV-RKSTCRSICSGSQDERRSVKQAQDEDPVLSFWGILKLNITWEPYVVGILCA 716

QY 720 IINGLQPAFSAIFSRIGIFTRDEDPETKRONSMFVFLVLGILSIFTFIFLQGTFFG 779

DB 717 VINGCMQPVFSIVFGIGVTRDDPKTKQCNKLFSLFFLVGMGICFVYFQGTFFG 776

QY 780 KAGEILTKRLRYMFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGAIGSLAVITQ 839

DB 777 KAGEILTKRLRYMFKMLRQDISWFDHNRNSTGALTTRLASDAANVKGAMSSLAGITQ 836

QY 840 NTANLGTGIIISLYGWQLTLLLLAIVPIAIVGVEKMKLSGQALKDKKELEGAKIAT 899

DB 837 NVANLGTGIIISLYGWQLTLLLVVVIAPLIILSGMMENKVLGQALKDKKELEVSIGKIAT 896


```
Qy 658 DSGSLIKRRSTRRSIIHAPQDQDRKLGTKEDLNENVPVSFWRLKLNSTWEPYVVGIF 717
Db 656 ESKSPLI-RRSIRSIHRRQDERRLSKEDVDVPMVFWQLKLNISEWPLVVGVL 714
Qy 718 CAIINGLOPAFISIFRSIGITFRDDEPTKRONSMFSLFVLVGLGIISFIFFLQGF 777
Db 715 CAVINGCIQPVFAIVFKSIGVFSRDDHETKORCNLFSLFLVMGMISFVYFFQGT 774
Qy 778 FGAGEILTKRLRYMFRSMRLQDVSFDDPKNTGALTTRLANDAAQVKGATGSLAVI 837
Db 775 FGAGEILTKRLRYMFRSMRLQDVSFDDPKNTGALTTRLANDAAQVKGATGSLAVI 834
Qy 838 TONTANLGTGIIS--LIYQWQLTLLLLAIVPIAIVGVVEMKMSQALKDKKELEG 895
Db 835 TONVANLGTGIISLVLVYQWQLTLLLVIIPLVLGSIEMKLLSQALKDKKELEISG 894
Qy 896 KIATEATENRTVSLTREOKFEYMAQSLQVYPYRNSLRKAHIFGVFSFTQAMMYFSYA 955
Db 895 KIATEATENRTVSLTREOKFEYMAQSLQVYPYRNSLRKAHIFGVFSFTQAMMYFSYA 954
Qy 956 GCFRFGAYLVANEFMNFODVLLVFSATVFGAMAVGVSSFPADYAKAKYSAHVIMIEK 1015
Db 955 ACFRGAYLVARELMTFENVWLVFSVAVFGAAGNTSSFPADYAKAKYSAHIIIGIEK 1014
Qy 1016 SPLIDSVPHGLKPTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKKGOTLALVSSGC 1075
Db 1015 IPEIDSVTEGLKPNWLEGNVKNVKNYPTRPNIPVLQGLSLEVKKGOTLALVSSGC 1074
Qy 1076 GKSTVOLLERFVPLAGSLVDGKEIKHLNVQWLRHAHLGIVSQEPILFDCSTAEINAYG 1135
Db 1075 GKSTVOLLERFVPMAGTVFLDGKEIKQLNVQCVRA-LGIYSQEPILFDCSTAEINAYG 1133
Qy 1136 DNSRVYSHEEIMOAAKEANIIHFIETLPEKYNTRVGDGTQSLGGQKQRTATARALVROP 1195
Db 1134 DNSRVYSHEEIVAREANIIHFIETLPEKYNTRVGDGTQSLGGQKQRTATARALVROP 1193
Qy 1196 HILLDEATSALDTESEKVVQVQALDKAREGRTCVIAHRLSTIONADLIVVFQNGVKKEH 1255
Db 1194 HILLDEATSALDTESEKVVQVQALDKAREGRTCVIAHRLSTIONADLIVVQNGVKKEH 1253
Qy 1256 GTHQQLLAQGIYFISMISVOAGAKR 1280
Db 1254 GTHQQLLAQGIYFSM--VOAGAKR 1276

RESULT 7
MDR3_HUMAN
ID MDR3_HUMAN STANDARD; PRT; 1279 AA.
AC P21439;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (P-glycoprotein 3).
GN ABCB4 OR PGY3 OR MDR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=89138016; PubMed=2906314;
RA van der Bliek A.M., Koolman P.M., Schneider C., Borst P.;
RT "Sequence of mdr3 cDNA encoding a human P-glycoprotein.";
RL Gene 71:401-411(1988).
RN [2]
RP MEDLINE=8811519; PubMed=2892668;
RA van der Bliek A.M., Baas F., ten Houte de Lange T., Koolman P.M.,
RT "The human mdr3 gene encodes a novel P-glycoprotein homologue and
RT gives rise to alternatively spliced mRNAs in liver.";
RL EMBO J. 6:3325-3331(1987).
```

```

RN [3]
RP GENE STRUCTURE.
RX MEDLINE=91161629; PubMed=2002063;
RA Lincke C.R., Smit J.J.M., van der Velde-Koerts T., Borst P.;
RT "Structure of the human MDR3 gene and physical mapping of the human
MDR locus.";
RL J. Biol. Chem. 266:5303-5310(1991).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS. HUMAN MDR3 IS NOT
CAPABLE OF CONFERRING DRUG RESISTANCE. MEDIATES THE TRANSLOCATION
OF PHOSPHATIDYLCHOLINE ACROSS THE CANALICULAR MEMBRANE OF THE
HEPATOCTYTE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- DISEASE: DEFECTS IN ABCB4 ARE A CAUSE OF PROGRESSIVE FAMILIAL
INTRAHEPATIC CHOLESTASIS TYPE III (PFIC), A FORM OF AUTOSOMAL
RECESSIVE LIVER DISORDERS, CHARACTERIZED BY EARLY ONSET OF
CHOLESTASIS THAT PROGRESSES TO CIRRHOSIS AND LIVER FAILURE BEFORE
ADULTHOOD.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; M23234; AAA36207.1; -.
CC EMBL; X06181; CAA29547.1; -.
CC PIR; J50051; DVHU3.
CC PIR; A42213; A42213.
CC HSSP; P13569; INBD.
CC MIM; 171060; -.
CC MIM; 602347; -.
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001140; ABC_transporter_tmem.
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR001687; ATP-GTP-A.
CC Pfam; PF00664; ABC_membrane; 2.
CC Pfam; PF00005; ABC_tran; 2.
CC SMART; SM00382; AAA; 2.
CC PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 192 211 POTENTIAL.
FT TRANSMEM 216 235 POTENTIAL.
FT TRANSMEM 301 320 POTENTIAL.
FT TRANSMEM 336 354 POTENTIAL.
FT DOMAIN 355 711 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 712 732 POTENTIAL.
FT TRANSMEM 756 776 POTENTIAL.
FT TRANSMEM 832 851 POTENTIAL.
FT TRANSMEM 854 873 POTENTIAL.
FT TRANSMEM 937 956 POTENTIAL.
FT TRANSMEM 976 993 POTENTIAL.
FT DOMAIN 994 1279 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 429 436 ATP (BY SIMILARITY).
FT NP_BIND 1069 1076 ATP (BY SIMILARITY).
FT REPEAT 1 640
FT REPEAT 641 1279
FT CARBOHYD 91 91
FT CARBOHYD 97 97
FT CONFLICT 1093 1093
SQ SEQUENCE 1279 AA; 140682 MW; 3D58C98B5C8D6087 CRC64;
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
V -> VFVDFGFO (IN REF. 2).
Query Match 75.7%; Score 4898; DB 1; Length 1279;
Best Local Similarity 75.4%; Pred. No. 1.2e-269;
Matches 971; Conservative 134; Mismatches 163; Indels 20; Gaps 9;
```

QY	1	MDPEGRKG-----SAENFWKMGKSKKKEKKKPT--VSTAMFRYSNNKLDLYMLV	52
Db	1	MDLEAKNGTAWRPTSAEGDF-ELGISKKRKTKTVMKGVTLFLFRYSQWQKLFMSL	59
QY	53	GTMALIHGAALPLMLVFGNMTDSFANAGTSRNTKTPVVIINESTTNTQHFINHLEEM	112
Db	60	GTIMAHSGGGLPLMLVFGEMTDFVDT--AGNFSFPNPSLSLLNPKG---ILEEM	113
QY	113	TTYAYIYSGIGAGVLVAAYIOVSFWCLAAGROILKIRKOFFHAIMRQEIFGWFVDHVGEL	172
Db	114	TRYAYIYSGIGAGVLVAAYIOVSFWTLAGROIRKIROKFFHAILRQEIFGWFINDITTEL	173
QY	173	NTRLTDDVSKINEGIDKVGCMFFQSTAFPTFTGFIYFPGWKLTVILAIISPVLGLSAAI	232
Db	174	NTRLTDDISKISEGIDKVGCMFFQAVATFAGFIYGFIRGWKLTIVIMAIISPILGLSAAV	233
QY	233	WAKILSFTDKELLAYAKAGAAVEELAAIRTIVIAFGQKKELELYNKNLEAKRIGIKK	292
Db	234	WAKILSFTDKELLAYAKAGAAVEELAAIRTIVIAFGQKKELELYNKNLEAKRIGIKK	293
QY	293	ATANISGAALFLIYASALAFVTGTSVLSEYTGIVGLTVFVFFSLIGAFSIGQASPS	352
Db	294	AISANISMGITAFLLIYASALAFVTGTSVLSEYTGIVGLTVFVFFSLIGAFSIGQASPS	353
QY	353	IEAFANARGAAYEIRKIIDNKPSIDSYSGHKPNKGNLEKFNHFSYPSRANVKILK	412
Db	354	IDAFANARGAAYEIRKIIDNKPSIDSYSGHKPNKGNLEKFNHFSYPSRANVKILK	413
QY	413	GLNLKVSQGTVALVGNSSCGKSTTVOLMQRLYDPDPGMVCDIGQDRTINVRHLREITG	472
Db	414	GLNLKVSQGTVALVGNSSCGKSTTVOLMQRLYDPDPGMVCDIGQDRTINVRHLREITG	473
QY	473	VVSQEPVLFATTAENIRVGRNWTDEIEKAVKANAYDFIMKLPNKEDTLVGERGARL	532
Db	474	VVSQEPVLFATTAENIRVGRNWTDEIEKAVKANAYDFIMKLPNKEDTLVGERGARL	533
QY	533	SGGOKORIAALVRNPKILLDEATSAIDTSEAVQVVALDKARKGRTTIVIAHRLST	592
Db	534	SGGOKORIAALVRNPKILLDEATSAIDTSEAVQVVALDKARKGRTTIVIAHRLST	593
QY	593	VRNADVTAGDDGVIVVEKGNHDEMEKEGIYFKLVTMQRTGNIELEENATGSKESDAL	652
Db	594	VRNADVTAGDDGVIVVEKGNHDEMEKEGIYFKLVTMQRTGNIELEENATGSKESDAL	651
QY	653	EMSPKSGSSILKRRSPRRSHIAPQODRKLGTKEG--LNEVPSVSVFWRLLKLNSEWPI	711
Db	652	EMSPKSGSSILKRRSPRRSHIAPQODRKLGTKEG--LNEVPSVSVFWRLLKLNSEWPI	710
QY	712	FVVGIFCAIINGGLQAPAFSIFRSIIGIFTRDEDPETKRONSMNFSVLFLVLGIIISFTF	771
Db	711	FVVGIFCAIINGGLQAPAFSIFRSIIGIFTRDEDPETKRONSMNFSVLFLVLGIIISFTF	769
QY	772	FLOGFTFKAGEILTKRLRYVFRSMLRDVDFDOPKNTTGALTFRLANDAAQVKGALG	831
Db	770	FLOGFTFKAGEILTKRLRYVFRSMLRDVDFDOPKNTTGALTFRLANDAAQVKGALG	829
QY	832	SRLAVITONTANLGTGIIISFIYQWLTLLLAIVPIIAITAGVEMKLSGQALKDKKEL	891
Db	830	SRLAVITONTANLGTGIIISFIYQWLTLLLAIVPIIAITAGVEMKLSGQALKDKKEL	889
QY	892	EGAKIATEAENRPTVSTREOKFEYMAQSLQVDPYRNSLRKAHFGVSPITQAMMY	951
Db	890	EGAKIATEAENRPTVSTREOKFEYMAQSLQVDPYRNSLRKAHFGVSPITQAMMY	949
QY	952	FSYAGCFRFGAYLVANEFMFDVLLVFSALVFGAMAVGVSSFAPDYAKAKYSAAHVIM	1011
Db	950	FSYAGCFRFGAYLVANEFMFDVLLVFSALVFGAMAVGVSSFAPDYAKAKYSAAHVIM	1009
QY	1012	IIKSPILDSYSPHGLKPTNLEGVNTVEVFNPTRPDIIVLOGLSLEKVKQGTALVG	1071
Db	1010	LFEROPLTDSYSEGLKPKDEGNITNEVFNPTRPDIIVLOGLSLEKVKQGTALVG	1069
QY	1072	SSGCKSKTVQVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRAGHGIYVSEPIFLDCSIAEN	1131
Db	1070	SSGCKSKTVQVOLLERFYDPLAGSVLIDGKEIKHLNVQWLRAGHGIYVSEPIFLDCSIAEN	1129
QY	1132	IAYGDSNRVSHVEIMOAAKEANHHFIETLPEKYNTRVGDGKTQLSGGOKORIAAL	1191
Db	1130	IAYGDSNRVSHVEIMOAAKEANHHFIETLPEKYNTRVGDGKTQLSGGOKORIAAL	1189
QY	1192	VRQPHILLDEATSAIDTSEKVVQVQALDKARERTCIVIAHRLSTIONADLIWVFQNGK	1251
Db	1190	VRQPHILLDEATSAIDTSEKVVQVQALDKARERTCIVIAHRLSTIONADLIWVFQNGK	1249
QY	1252	VKEHGTQHOQLLAQKGIYFMSVQAGAK 1279	
Db	1250	VKEHGTQHOQLLAQKGIYFMSVQAGTQ 1277	
RESULT 8			
MDR2_MOUSE			
ID	MDR2_MOUSE	STANDARD;	PRT; 1276 AA.
AC	P21440;		
DT	01-MAY-1991 (Rel. 18, Created)		
DT	01-MAY-1991 (Rel. 18, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Multidrug resistance protein 2 (P-glycoprotein 2).		
GN	PGY2 OR PGY-2 OR MDR2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
FX	MEDLINE=88302195; PubMed=3405218;		
RA	Gros P., Raymond M., Bell J., Housman D.;		
RT	"Cloning and characterization of a second member of the mouse mdr		
RT	gene family.";		
RL	Mol. Cell. Biol. 8:2770-2778(1988).		
RN	[2]		
RP	SEQUENCE OF 1-23 FROM N.A.		
RC	STRAIN=BALE/C;		
RA	Kirschner L.S., Horwitz S.B.;		
RL	Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.		
CC	-!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED		
CC	DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS. MOUSE MDR2 IS		
CC	NOT CAPABLE OF CONFERRING DRUG RESISTANCE.		
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein.		
CC	-!- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE		
CC	RELATED BUT DISTINCT CELLULAR GENES.		
CC	-!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; J03398; AAA39516.1; -		
DR	EMBL; M74151; AAA39515.1; -		
DR	PIR; A30409; DVMS2.		
DR	HSSP; P13569; INBD.		
DR	MGI; MGI:97569; Pgy2.		
DR	InterPro; IPR003593; AAA.		
DR	InterPro; IPR001140; ABC_transporter_tmem.		
DR	InterPro; IPR003439; ABC_transporter.		
DR	InterPro; IPR001687; ATP_GTP_A.		
DR	Pfam; PF00564; ABC_membrane; 2.		
DR	Pfam; PF00005; ABC_tran; 2.		
DR	SMART; SM00382; AAA; 2.		
DR	PROSITE; PS00211; ABC_TRANSPORTER; 2.		
DR	ATP-binding; glycoprotein; Transmembrane; Transport; Repeat;		
KW	Multigene family.		
KW	DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).		
FT			


```
FT TRANSMEM 52 75 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT DOMAIN 346 708 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 708 728 POTENTIAL.
FT TRANSMEM 753 773 POTENTIAL.
FT TRANSMEM 829 849 POTENTIAL.
FT TRANSMEM 850 870 POTENTIAL.
FT TRANSMEM 933 950 POTENTIAL.
FT TRANSMEM 970 990 POTENTIAL.
FT DOMAIN 991 1276 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (BY SIMILARITY).
FT NP_BIND 1066 1073 ATP (BY SIMILARITY).
FT REPEAT 1 635
FT REPEAT 636 1276
SQ SEQUENCE 1276 AA; 140332 MW; A6C38DCD2C118EEF CRC64;

Query Match 75.0%; Score 4850.5; DB 1; Length 1276;
Best Local Similarity 74.6%; Pred. No. 5.7e-267;
Matches 958; Conservative 140; Mismatches 171; Indels 15; Gaps 8;

QY 1 MDPEGKRGSAEK--NFWKMGKSK--KEKKEKKPTVSTFAMFRYSNWLRLYMLVGTMA 56
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 1 MDLEAARNGTARRLDGDFELGSIISNOGREKKKKVNLIGLLTLFRYSQWQDKLNFELGTLM 60

QY 57 AIIHGAALPLMLVFGNMTDSFA--NAGISRNKTFPPVIINESINNTQHTNHLSEEMTTY 115
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 61 AIHAGSGPLMLVFGNMTDKFDVNTG---NFSLPVNFSLMNLNPGR---ILEEEMTRY 113

QY 116 AYYYSIGAGVLYAAYIQVSWFCLAGROILKIRKOPFHAIMRQEIQWDFVDVHVGELNTR 175
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 114 AYYYSGLGGVLYAAYIQVSEFWLLAGROILKIRKOPFHAILRQEWGDFDKIGTTELNR 173

QY 176 LTDDVSKINGIGDKVGMFTQSIATFTPTGTVGTPGWKLTLYLAISPVGLSAAIWA 235
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 174 LTDDVSKISEGIGDKVGMFTQSIATFTPTGTVGTPGWKLTLYLAISPVGLSAAIWA 233

QY 236 ILSFTDKELLAYAKAGAAVEELAAIRTVIAFGGOKKELERYKNLEAKRIGIKKAIT 295
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 234 ILSFTSKELAAAYAKAGAAVEELAAIRTVIAFGGOKKELERYKNLEAKRIGIKKAIT 293

QY 296 ANISGAFFLLIYASVALAFWYGTSLVSLSEYITIGOVILVFFSVLIGAFSIGOASPSIDA 355
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 294 ANISMGIAFLIYASVALAFWYGTSLVSLSEYITIGOVILVFFSVLIGAFSIGOASPSIDA 353

QY 356 FANARGAAEYFKLIIDNKPSIDYSKSGHPDNKGNLFBNKVFHPSRKVEYKILKGLN 415
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 354 FANARGAAEYFKLIIDNNPKIDFSERGHKPDNIKGNLFSDVHFSYPSRANIKILKGLN 413

QY 416 LKVSOGTVALVNSGCGKSTTVOLMORLYDPTDGMVCIDGQDIRTNVHRLREITGVVYS 475
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 414 LKVSOGTVALVNSGCGKSTTVOLMORLYDPTDGMVCIDGQDIRTNVHRLREITGVVYS 473

QY 476 QEPVLFPATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNFKFTDLVGERGALSG 535
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 474 QEPVLFPATTIAENIRYGRNVMTDEIEKAVKEANAYDFIMKLPNFKFTDLVGERGALSG 533

QY 536 OKORIAIARALVNPRIKILLDEATSDLTDESEAVVOVALDKARKGRTTIVIAHRLSTVRN 595
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 534 OKORIAIARALVNPRIKILLDEATSDLTDESEAVVOVALDKARKGRTTIVIAHRLSTVRN 593

QY 596 ADVIAGFDGVIYKGNHDELKKEGIFYKLVTMOPRGNEIELENATGESKESDALEMS 655
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 594 ADVIAGFDGVIYKGNHDELKKEGIFYKLVTMOPRGNEIELENATGESKESDALEMS 653

QY 656 PKDSGSLIKRRTRRSIHAPOQDKRLGTKEDLNENVPVSFWRIKLNSTEWPFVYVG 715
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 654 P-NGWKARIFRNSTKSLKSPH-QNRLDEETNELDANVPVFLVKLNKTNWPFVYVG 711

QY 716 IFCAILINGLOPAFSIIFRSRIIGFTIRDEDPETKQNSNMFSVLFLVLVIISFITFFLQ 775
```

```
Db 712 TVCAIANGALQPAFSIILSEMIATFGPGDD-AVKQKCNMFSVLVLGLVLSFFTFFLQ 770
QY 776 FTFGKAGEILTTLRLYRVFMSMLRQDVSWFDDPKNTGTALTTRLANDAAQVKAIGSRILA 835
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 771 FTFGKAGEILTTLRLSRMAFKAMLRQDMSWFDHKNSTGALSTRLATDAQVQCATGKLA 830
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 836 VITQNTANLGTGIIISLIYQWLTLLLLAIVIPIIATAGVVMKLSGQALKKKKELEGAG 895
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 831 LIAQNTANLGTGIIISFIYQWLTLLLSVVPPIAVAGIVEMKLAGNAKDKKEMEAG 890
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 896 KIATEAENPRTVTVSLTREOKFEYMAQSLQVPRNSLRKAHIFGVYSFSTQAMMYFSYA 955
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 891 KIATEAENPRTVTVSLTREOKFEYMAQSLQVPRNSLRKAHIFGVYSFSTQAMMYFSYA 950
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 956 GCFRFGAYLVANEFMNFQDVLVLSAIVFGAMAVGVSSFAPDYAKAKYSAAHVIMIEK 1015
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 951 GCFRFGSILYVNGHMRFKDVLVLSAIVFGAMAVGVSSFAPDYAKAKYSAAHVIMIEK 1010
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 1016 SPLIDSYSPHGLKPNLTLEGNTVFNEVVPTRPDIPVLOGLSLEVKKQOTLALVSSGC 1075
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1011 QPLIDSYSSEGLWPKDFEGSVTFNEVVPTRANVPVLOGLSLEVKKQOTLALVSSGC 1070
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 1076 GKSTVQVLLERFVDPLAGSLVDGKEIKHLNQLWLRHGLIVSQEPILFDCSIAENIAG 1135
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1071 GKSTVQVLLERFVDPLAGSLVDGKEIKHLNQLWLRHGLIVSQEPILFDCSIAENIAG 1130
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 1136 DNSRVVSHBEIMOAAKEANIHFIETPLPEKYNTRYVGDGKGTQSLGGQKORIAIARALVRP 1195
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1131 DNSRVVPHDEIVRAAKEANIHFIETPLPEKYNTRYVGDGKGTQSLGGQKORIAIARALVRP 1190
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 1196 HILLDEATSDLTDESEKVVQVQALDKARERTCIVIAHRLSTIONADLIIVFQNGKVEH 1255
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1191 RVLLDEATSDLTDESEKVVQVQALDKARERTCIVIAHRLSTIONADLIIVFQNGKVEH 1250
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 1256 GTHQQLLAQGIYFMSISVQAGAK 1279
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1251 GTHQQLLAQGIYFMSVNIQAGTQ 1274
  ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

RESULT 9
MDR3_CRIGR
ID MDR3_CRIGR STANDARD; PRT; 1281 AA.
AC P23174;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (P-glycoprotein 3).
GN PGY3 OR PGY3.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RP [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=92135896; PubMed=1685679;
RA Endicott J.A., Sarangi F., Ling V.;
RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
  gene family.";
RL DNA Seq. 2:89-101(1991).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
  DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT
  CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
  CANNOT.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M60042; AAA68885.1; -
DR HSP; P13569; 1NBD.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00564; ABC_membrane; 2.
DR Pfam; PF00005; ABC_membrane; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; P50021; ABC_TRANSPORTER; 1.
KW ATP-binding; glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 57
FT TRANSMEM 58 78
FT TRANSMEM 122 142
FT TRANSMEM 191 211
FT TRANSMEM 218 238
FT TRANSMEM 299 319
FT TRANSMEM 328 348
FT DOMAIN 349 712
FT TRANSMEM 713 733
FT TRANSMEM 758 778
FT TRANSMEM 834 854
FT TRANSMEM 855 875
FT TRANSMEM 938 958
FT TRANSMEM 975 995
FT DOMAIN 996 1281
FT NP_BIND 429 436
FT NP_BIND 1071 1078
FT SEQUENCE 1281 AA; 140866 MW; 2203EF61EBB29602 CRC64;

Query Match 74.4%; Score 4810; DB 1; Length 1281;
Best Local Similarity 73.5%; Pred. No. 1.1e-264;
Matches 947; Conservative 148; Mismatches 175; Indels 18; Gaps 8;

QY 1 MDEGGKGSAAEK-----NFKMGKSK--KEKKKKPTVSTFAMPYRNWLDRLMLVG 53
DB 1 MDLEAARNGTARPGTVGDFEUGSISNGRNKKKKVNLGPTLFRYSDWQDKLFMLIG 60
QY 54 TMAAIIHGAALPLMLVGNMDSFA-NAGISRNKTFPVIINESITNTQHTFNHLEEM 112
DB 61 TMAIAHSGSLPLMIVFGEMTDKFNAG---NFSLPVNFSLMINPGR---ILEEM 113
QY 113 TTYAYTSGAGVLAAYIQVSWFLAAGROILKIRKOFFHAIMRQETGWFDVHDVDEL 172
DB 114 TRYAYTSGLGGVLAAYIQVSWFLAAGROILKIRKOFFHAIMRQETGWFDVHDVDEL 173
QY 173 NTRLTDDVSKINIGIGDKVMFTQSTATFTTGTGVTGPKWKLTLVILATSPVLGLSAAI 232
DB 174 NTRLTDDISKISGIGDKVMFTQSTATFTTGTGVTGPKWKLTLVILATSPVLGLSAAV 233
QY 233 WAKILSFTDKELLAYAKAGAAVEVLAAIRTVIAFGQKKELERYNKNLEAKRIGIKK 292
DB 234 WAKILSFTDKELAAAYAKAGAAVEVLAAIRTVIAFGQKKELERYNKNLEAKRIGIKK 293
QY 293 AITANISIGAFLLIYASALAFWYGTSLVLSSEYITIGQVLTFFSVLIGAFSIGOAPS 352
DB 294 AISANISGIAFLLIYASALAFWYGTSLVLSSEYITIGQVLTFFSVLIGAFSIGOAPS 353
QY 353 IEAFANARGAAEYFKIIDNKPDSISYSGSKPKDNKGNLEKFNHFSYPSRKEYKILK 412
DB 354 IDAFANARGAAEYFKIIDNKPDSISYSGSKPKDNKGNLEKFNHFSYPSRKEYKILK 413
QY 413 GLNLKVGSGQTVALVNGSGCGKSTTVQLMQRLYDPTDGVICDQDQIRTNVRLREITG 472
DB 414 GLNLKVGSGQTVALVNGSGCGKSTTVQLMQRLYDPTDGVICDQDQIRTNVRLREITG 473
QY 473 VVSEPVLFATIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPKPFDTLVGERGAL 532
DB 474 VVSEPVLFATIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPKPFDTLVGERGAL 533

DB 474 VVSEPVLFATIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPKPFDTLVGERGAL 533
QY 533 SGGQKQRIATARALVRNPKILLDEATSDALDTESEAVVQVALDKARKGRITVIAHRLST 592
DB 534 SGGQKQRIATARALVRNPKILLDEATSDALDTESEAVQAAALDKARKGRITVIAHRLST 593
QY 593 VRNADVIAGDDGVIVKEGNHDELMKEGIFYFLKVTMOTRGNIELEENATGESKESDAL 652
DB 594 VRNADVIAGDDGVIVKEGNHDELMKEGIFYFLKVTMOTRGNIELEENATGESKESDAL 653
QY 653 EMSPKDSGSSILIKRSTRSRSHIAPQGDRLKGTED--LNEVPSVSFWRLIKLNSTEWPY 711
DB 654 GMP-NGWKSHPFRNSTKSLKSRHHRLDVEDDDELANDVPVSEFLKVLKLNKTEWY 712
QY 712 FVVGIFCAIINGGLQPAFISFIISRIIGIFTRDEPETKQNSMFSVFLVGLGIISFITF 771
DB 713 FVVGTCVAINGALQPAISIIISEMIAIFGPGD-AVKQCKNLFSVFLVGLGLVLSFTF 771
QY 772 FLOGFTFKAGEILTKRLRYMFRSMRLQDVSFDDPKNTGTALTTLANDAAQVKAIG 831
DB 772 FLOGFTFKAGEILTKRLRYMFRSMRLQDVSFDDPKNTGTALTTLANDAAQVKAIG 831
QY 832 SRLAVITONTANLGTIIISLYGWQLTLLLAIVPIIAIAGVVMKMLSGQALKDKKEL 891
DB 832 TRALIAQNTANLGTIIISLYGWQLTLLLAIVPIIAIAGVVMKMLSGQALKDKKEL 891
QY 892 EGAGKIATEALENFRVTYVSLTRQKFEYMAQSLQVYPYRNSLRKAHIFGVFSITQAMMY 951
DB 892 EAAGKIATEALENFRVTYVSLTRQKFEYMAQSLQVYPYRNSLRKAHIFGVFSITQAMMY 951
QY 952 PSYAGCFRFGAYLVANEFMNFQDVLVFSATVFCGAMAGQVSSFPADYAKAKYSAHVIM 1011
DB 952 PSYAGCFRFGAYLVANEFMNFQDVLVFSATVFCGAMAGQVSSFPADYAKAKYSAHVIM 1011
QY 1012 IIEKPLDYSYSPHGLKNTLEGNVTNEVFNTPRPDIPVLOQLSLEVKKGOTLALVG 1071
DB 1012 LFEKPLDYSYSGELWPKDFEGSVTFNEVFNTPRANMPVLQGLSLEVKKGOTLALVG 1071
QY 1072 SSGCGKSTVQVLLERFYDPLAGSLIDGKEIKHLNVOVLAHILGIVSQEPILFPCSTAE 1131
DB 1072 SSGCGKSTVQVLLERFYDPMAGTVLLDQEAQKLNIOVLAHILGIVSQEPILFPCSTAE 1131
QY 1132 TAYGNSRVVSHETMOAAKEANIHHTETLPEKYNTRVGDKGTQSGQKQRIATARAL 1191
DB 1132 TAYGNSRVVSHETMOAAKEANIHHTETLPEKYNTRVGDKGTQSGQKQRIATARAL 1191
QY 1192 VRQPHILLDEATSDALDTESEKVVQEAALDKARKGRITVIAHRLSTQNADLIYVFQNGK 1251
DB 1192 VRQPHILLDEATSDALDTESEKVVQEAALDKARKGRITVIAHRLSTQNADLIYVFQNGK 1251
QY 1252 VKEHGTQOQLLAQKGIYFSMISVQAGAK 1279
DB 1252 VKEHGTQOQLLAQKGIYFSMISVQAGAK 1279

RESULT 10
MDR2_RAT ID MDR2_RAT STANDARD; PRT; 1278 AA.
AC Q08201;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 2 (p-glycoprotein 2).
GN PGY2 OR MDR2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISCHER;
RX MEDLINE=93376516; PubMed=8103593;
RA Brown P.C., Thorgerirsson S.S., Silverman J.A.;

"Cloning and regulation of the rat mdr2 gene.";
Nucleic Acids Res. 21:3885-3891(1993).
-!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: L15079; AAA02937.1; -;
DR HSP: P13569; INBD.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC_transporter_tmem.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane; 2.
DR Pfam: PF00005; ABC_tran; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 54 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 55 75 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT DOMAIN 346 709 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 710 730 POTENTIAL.
FT TRANSMEM 755 775 POTENTIAL.
FT TRANSMEM 831 851 POTENTIAL.
FT TRANSMEM 852 872 POTENTIAL.
FT TRANSMEM 935 955 POTENTIAL.
FT TRANSMEM 972 992 POTENTIAL.
FT DOMAIN 993 1278 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (POTENTIAL).
FT NP_BIND 1068 1075 ATP (POTENTIAL).
SQ SEQUENCE 1278 AA; 140655 MW; D4FB6BE745AF73BF CRC64;

Query Match 74.2%; Score 4796.5; DB 1; Length 1278;
Best Local Similarity 73.7%; Pred. No. 6.5e-264;
Matches 947; Conservative 146; Mismatches 177; Indels 15; Gaps 8;

QY 1 MDPGGRGSAEK--NFWKMGKKS--KKEKKEKPTVSTFAMFRYSNNWLDRLYLVTGMA 56
DB 1 MDLEARNGTARRLDGDFELGSLISNQSREKKKVNLIPTLFYSDWQDKLFLMLIGTAM 60
QY 57 AITHGAALPLMLVFGNWDSPA-NAGISRNKTPFVINESTNNQHPINLHEEMTY 115
DB 61 AIAHSGSLPLMLVFGEMTDKFDVNDAG--NFSLPVNFSLMLNPGR----LLEEMTRY 113
QY 116 AYTYSIGAGVLAAYIQVSWFLAAGROILKIRKOFFHAIMROETGWDFVDHVGELNTR 175
DB 114 AYTYSIGLGGVLLAAYIQVSWFLAAGROIRKIRKFFHAILRDQEMGFDKGTTELNTR 173
QY 176 LTDVSKINGIDKVGMEFFQSFIATFFTGFTPGWKLTLVLAISPLVGLSAAIAWAK 235
DB 174 LTDIDSKISBGIDKVGMEFFQAIATFFAGFIVGFRGNKLTIVINAILGLSTAWAK 233
QY 236 ILSFDTDKELLAYAKAGAAVEALAIRTVIAFGQKKLELYRNKNLEAKRIGIKKAIT 295
DB 234 ILSFSDKELAAAYAKAGAAVEALAIRTVIAFGQKKLELYRNKNLEAKRIGIKKAIT 293
QY 296 ANISGAIFLLIYASVALAFWGTSLVLSSEYTIQVLTVFESVLIGAFSGQASPSIEA 355
DB 294 ANISMGIAFLLIYASVALAFWGTSLVLSSEYTIQVLTVFESVLIGAFSGQAPCIDA 353

QY 356 FANARGAAVEIPKIIDNKPSIDSYSKSGHKPDNIKGNLEFNHVSFSPSRKEVKILKGLN 415
DB 354 FPNARGAAVVIDIIDNNPKIDSFSEHGHPDSIKGNLEFSDVHVSFSPSRANIKILKGLN 413
QY 416 LKVGSGQTVALVNGSGCKSTTVOLMORLYDPTDGMVCIQDQDITRTIVNHLREITGVVS 475
DB 414 LKVGSGQTVALVNGSGCKSTTVOLQRLYDPTDGTISIDQDINRNFVRLREFIGVVS 473
QY 476 QEPVLFAITIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPKFDTLVGERGARLSGG 535
DB 474 QEPVLFTTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPKFDTLVGERGARLSGG 533
QY 536 QKQRTAIRALVRNPKILLDEATSEAVVVALDKARKGRTTIVIAHRLSTVRN 595
DB 534 QKQRTAIRALVRNPKILLDEATSEAVVVALDKARKGRTTIVIAHRLSTVRN 593
QY 596 ADVIAGFDGIVVEKGNHDELMKEGIVFKLVMTQRTGNETELENAETESKESDALEMS 655
DB 594 ADVIAGFDGIVVEGSHSELIKKEGIVFRLVNMQTSQSLSEEFVEVLSDEKAAGGVA 653
QY 656 PRDGSGLIKRRSTRRSIHAPQGDRLGKTK-EDLNENVPVSFVWRILKLNSTEWPFV 714
DB 654 P-NGWKARIFERNSTKSLKSSRAHQNRDLDETNELDANVPVSFLVRLNKTWPFV 712
QY 715 GIFCAILINGLOPAFSIIFSRIGIFTRDEDPETKRONSNFVFLVLGLIISFITFFLQ 774
DB 713 GTLCAIANGLOPAFSIILSEMIATIFGPGDD-TVYQKCNMFSLVFLGLGVHFFTFLLQ 771
QY 775 GFTFGKAGEILTTLRLVWFRSMURQDVSWFDDPKNTGALTTRLANDAAVKGATGSR 834
DB 772 GFTFGKAGEILTTLRLVWFRSMURQDVSWFDDPKNTGALTTRLANDAAVKGATGSR 831
QY 835 AVITONTANLGTIIISLIYGMWLTLLLLAIVPIIAGVVEKMLSGQALDKKKELEGA 894
DB 832 ALIAQNTANLGTIIISLIYGMWLTLLLLAIVPIIAGVVEKMLSGQALDKKKELEGA 891
QY 895 GKIAEALENFRVTSLTREQKFEYMYAQSLOVPYRNSLRKAHIFGVFSITQAMMYFSY 954
DB 892 GKIAEALENFRVTSLTREQKFEYMYAQSLOVPYRNSLRKAHIFGVFSITQAMMYFSY 951
QY 955 AGCFRFGAYLVANEFMNFQDVLVFSVIAVFGAMAVGVSFADPYAKAKVSAHVIMITE 1014
DB 952 AGCFRFGAYLVANEFMNFQDVLVFSVIAVFGAMAVGVSFADPYAKAKVSAHVIMITE 1011
QY 1015 KSPILDSYSPHGLKPNLTLEGNTFNEVFNYPTRDPDIPVLQGLSLEVKKQTLALVSGS 1074
DB 1012 RQPLIDSYSREGMWPDKFEGSVTFNEVFNYPTRANVPVQLGLSLEVKKQTLALVSGS 1071
QY 1075 CGKSTVOLLERFVDPPLAGSLVDGKEIKHLNVOWLRAHLGIVSQBPILFDCSIAENTAY 1134
DB 1072 CGKSTVOLLERFVDPPLAGSLVDGKEIKHLNVOWLRAHLGIVSQBPILFDCSIAENTAY 1131
QY 1135 GDSRVVSHEEIMQAAKEANIHFETLPEKYNTRVGDGKTOLSGGQKORIAIARALVRO 1194
DB 1132 GDSRVVSQDEIVRAAKEANIHFETLPEKYNTRVGDGKTOLSGGQKORIAIARALVRO 1191
QY 1195 PHILLDEATSEKVVQVQALDKAREGRTTIVIAHRLSTVRN 1254
DB 1192 PRVLLDEATSEKVVQVQALDKAREGRTTIVIAHRLSTVRN 1251
QY 1255 HGTHQQLLAQKGIYFSMISVQAGAK 1279
DB 1252 HGTHQQLLAQKGIYFSMISVQAGAK 1276

RESULT 11
AB11 HUMAN
ID AB11 HUMAN STANDARD; PRT; 1321 AA.
AC Q95342; Q9UNB2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

Db	270	ELKAYAGSVADEVISSTRTVAAGGERKEVEREKNLVFAQRWGIKGIYWGFTGYM	329
Qy	304	FLLIYASYALAFWYGTSLVL-SSEYITIGVLTIVFFSVLIGAFSIGQASPSIEAFANARGA	362
Db	330	WCLIFCYALAFWYGLVLEEGEYSPGALQIFLSVLIIGALNLGNASCLCEAFAGRAA	389
Qy	363	AYEIFKIIDKNPKSIDYSKSGHKPDNIKLEPKNVHFSYPSRKEVKIILKGLNLKQVSOQ	422
Db	390	ASSIFETIDRPIIDCMSEDGYLERIKGEIEFHNFTYFSPRVEVKIILNLSWVTKPGE	449
Qy	423	TVALVNGSGCGKSTTVOLMORLVDPDTGMVCIDQDQIRITNVHRLREITGVSQEVPFLFA	482
Db	450	MTALVPGSGAGKSTALQILHRFFGPTGEGVTVESHDIRSHIQWLNRNQIGIVEQEPVLF	509
Qy	483	TTIAENIRYGRENVTWDEIEKAVKEANAYDFIMKLPNKFTDLVGERGARGLSGQKORIAI	542
Db	510	HTIAEKIRYGRREDATMEDILQAAKEANAYNFIMDLPOQFDTLVGEQGGMGQKQORVAI	569
Qy	543	ARALVRNPKILLDDEATSDALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADVIAG	602
Db	570	ARALIRNPKILLDMATSDALDNSEAMVQBALSKTQHGHTIVSAHRPATIRTDVILGC	629
Qy	603	DDGVIVKGNHDELMKEKGYFYKLVMTQTRGNETELE- ----NATGESKSESDALEMSPKD	658
Db	630	EHGRAVERGTEELLERKGYFVALTVLSQNRQDQENEDKATEDIDPEKTESRGNYQD	689
Qy	659	SGSLIKRRS- ----TRRSIHAPQGDQDKLGTCKEDNLNENVPVSFWRL	702
Db	690	SLRASLRQSKQLSYLAHEPPMAVEDHDKSTHEEDKDLPAQD- ----IEPASVRRIR	745
Qy	703	KLNSTEWPYFVVGTFCAINGLQPAFSIIFSRILGIFTRDEDPETKQKSNMFSVLFLV	762
Db	746	KLNAPEYMLLGSNGAANGAVTPIYAFLEFSQILGTFSL-PKQEQRSQINGICLLFVT	804
Qy	763	LGIISFTFFLQGFQFKAGBILTKRLRYMVFRSMLRQDVSWEDPKNTTGCALTTRLAND	822
Db	805	LGCVSFTFQFGVTFKAGSGELLTKRLKFGFRAMLQDQIGWFDLDRNSPGALTTRLATD	864
Qy	823	AAQVKGATGSLAVITONIANLNGIILSIYQWGLTLLAILVPIITAIAGVVMKMLSG	882
Db	865	ASQVQATGSGQIGMMVNTVNTVAMIIAFLFSKLLIGIVCFPPFLLASGALQTKMLTG	924
Qy	883	QALKDKKELEGAGKIATBAENFRTVSLTREQKPEYMYAQSLQVYRNSLRKAHIFGVS	942
Db	925	FASRDQALERAGQITSEALSNIRTVAGIGKERFIETFEAELEKPKYKMAIKKANVYGLC	984
Qy	943	FSITQAMVYFVYAGCFRGAIVLANEFMNFQDVLVLSAIVFGAMAVGVSSFAPDYAKA	1002
Db	985	FGFSQCITFIANSASYRGYGLISNEGHLHFSYFVRVISAIVLSATALGRASSYTPSTAKA	1044
Qy	1003	KVSAAHVIMITEKSLIDSYSPHGLKPNLTGNTVFNVEVFNYPTRPDIPVLOGLSLEVY	1062
Db	1045	KISAARFQLIDRQPPINVYSSAGEKWNFOGKIDFVDCCKEYPSRDPDIQVNLGLSYMS	1104
Qy	1063	KGQTLALVSSCGKSTVVQLLERYDPDLAGSVLIDGKEIKHLNVQMLRAHLGIVSQEPI	1122
Db	1105	PRQTLAFVSSCGKSTSIQLLERYDPDHGKVMIDHDSRKVNTQFLRSNIGIVSQEPV	1164
Qy	1123	LPDCSIAENIAGDNRVSVSHEEIMQAAKEANIHIFETLPEKYNTRVGDGTOLSGQK	1182
Db	1165	LPACSIKDNIKYDNTQELPMERIIIAAKAKQVHDFVMSLPEKETYNTVGSQSGLSRGEK	1224
Qy	1183	QRIATARLVROPHILLDDEATSDALDTESEKVVQBALDKAREGRTCIIVIAHRLSTIONAD	1242
Db	1225	QRIATARAIVRDPKILLDDEATSDALDTESEKTVQVALDKAREGRTCIIVIAHRLSTIQNSD	1284
Qy	1243	LIVFONGKVKHGHQHLQALQKGYIFYSMS	1273
Db	1285	IIAVASQGNVIEKGTHEELMVOKGAYKLV	1315

FT DOMAIN 777 794 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 795 815 POTENTIAL.
 FT DOMAIN 816 869 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 870 890 POTENTIAL.
 FT TRANSMEM 891 911 POTENTIAL.
 FT DOMAIN 912 979 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 980 1000 POTENTIAL.
 FT DOMAIN 1001 1011 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1012 1032 POTENTIAL.
 FT DOMAIN 1033 1321 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 455 462 ATP (POTENTIAL).
 FT NP_BIND 1113 1120 ATP (POTENTIAL).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 481 481 L -> P (IN REF. 2).
 FT CONFLICT 633 633 T -> V (IN REF. 2).
 SQ SEQUENCE 1321 AA; 146675 MW; 15B5EBF175D32967 CRC64;

Query Match 49.5%; Score 3201; DB 1; Length 1321;
 Best Local Similarity 49.6%; Pred. No. 1.5e-173;
 Matches 641; Conservative 236; Mismatches 378; Indels 38; Gaps 9;

QY 15 FWKMGKSKKKEK--KPTVSTFAMFRYSNWLRLMYLVTGMAAIIHGAALPLMLVFG 72
 DB 25 FHNDKKSRLQDKKKGARVGFELFRFSSKDNLMFMGSCALLHGMAGPMGIIVFG 84
 QY 73 NMTDSFANAGLSR-----NKTFPVI---INESTNNTQHPINLHEEMTYAYV 118
 DB 85 ILTDIFVEYDIKQELSGPKVCNMNTIVWINSFNQNTNGTSCGLVDINSEVIFKSGI 144
 QY 119 YSGIGAGLVAAAYIQSVFWCLAGROILKIRKOFFHAIMROETIGWDFVDHVGELNTRTD 178
 DB 145 YAGVGAVLILGYQIRLWITGAROIRKMKFYFRIMRMELGWFDCSTVSGELNSRSD 204
 QY 179 DVSKINGIGDKVGMFFQSTATPFTGIVGTPGWKLTLVILAISPLVGLSAAIWAIIKS 238
 DB 205 DIINKIDEAIDOMALFLQRLSTALSGLLGLFGYRGWKLTLVILASPLIGIAGAVIGLSVA 264
 QY 239 SFYDKELAYAKAGAAVEELAIRVIAFGGOKLELYKNLEAKRIGIKKAITANI 298
 DB 265 KFELELKAYAKAGSTADEVLSIRIVAAFGGENKEVEREYKLNMFQAQRWGIKGMVNGF 324
 QY 299 SIGAFLLIYASVALAFWTGTSVL--SSEYTIQCVLTFFSVLIGAFSTIGQASPSIEAFA 357
 DB 325 FTGYMCLIFFCVAFALAFWYGSRLVLDEGEYTFGLTIQIFLCVLIAMNIGNASSCLEIFS 384
 QY 358 NARGAAVEIFKIIDNKPSIDSYKSKHKPDNKGLEFNHVFSPSRKVKILKGLNLK 417
 DB 385 TGCSSAASSIFQITIDQPVMDCHMSGDGYKLDRIKGEIEFHNVTFFHPSRPEVKILNLSMV 444
 QY 418 VQSGOTVALVNGSGCKSTTVOLMORLYDPTDGMVCIDGQDRTINVRHLREITGVYSQE 477
 DB 445 IKGETTAFVSGSGAKSTALQIQRFYDPCGWTLOGHDIRSUNIRWLKDQIGIVEQE 504
 QY 478 PVLPTAIIENIRYGRNVTMDIEKAVEANAYDFIMKLPNKEDTLVGERGARGLSGQK 537
 DB 505 PVLFTTIAENIRLGRREATMEDIVQAARDANAYFIMALPQQPDTLVGEGGQMGSGQK 564
 QY 538 ORTAIRALVRPKILLDEATSDLTSEAVQVVALDKARKGRTTVIAHRLSTVRNAD 597
 DB 565 ORVAIRALIRPKILLDMATSDLDNESEAKVQOGLAKIQTGHTIISVAHRLSTVRAD 624
 QY 598 VIAGFDGVIVKGNHDELMKEGYFKLVMTQTRGNEIELENA--TGESKESDALEM-- 654
 DB 625 VIIGFPHGTAVERGTHEELLERKGYFMVLVLOSQEDNTHKETGKQDTEGTPERTF 684
 QY 655 ----SPKDGSSSLIKRSTR-----SIHAPQODRKLGTKEDLNENVPVSFW 699
 DB 685 SRGSYQDSLRSIRQSKSLSHLSHEPPALGDHKKSYEDRK--DNDVLVEEVPAPVR 742
 QY 700 RILKLNSTWPPYVVGIFCAIINGLQAFSIIIFRSIIGITFREDPETRQNSMFSVL 759

DB 743 RILKYNISEWPYILVGCALCAAINGAVTPIYSLLSFQILKTFSL--VDKEQQRSEIYSMCLF 801
 QY 760 FLVLGIISFITFFLQGTGKAGELLTKRLRYMVVFRSMRLQDYSVDPPKNTTCALTTRL 819
 DB 802 FVILGCVSLFTQFLOGYNFAKSGELLTKRKFGKRLKQDQIGWEDDLKNNPGVUTTRL 861
 QY 820 ANDAAQVKAIGSRLAVITQNIANLGTGIIISLIYQMLTLLLLAIVPIIAIAGVVMEMK 879
 DB 862 ATDASQVQCATGSQVGMVMSFTNIFAVLIAFLFNKLSLVISVFPPFLALSGAVQTKM 921
 QY 880 LSGQALDKKKELEGAGKATATENFTVYSLTRQKFEYMAQSLOVPYRNSLRKAHIF 939
 DB 922 LTGFASQDKLELEKAGQITNEALSNIETVAGIGVEGRFKAFKFEVELEKSYKTAIRKANY 981
 QY 940 GVFSITQAMMYFYAGCFRFGAYLVANEPNFDVLLVFSALVFCAMAVGOVSSFAPDY 999
 DB 982 GLCYAFSGQISFLANSAAVRYGGYLYVEDLNEYSYFVSVSSAMATAVGRTFSYTPSY 1041
 QY 1000 AKAKVSAHVMIETKSPIDSYSPLGKPKNTLEGNTNEVFNTPRPDIPLVQLGSL 1059
 DB 1042 AKAKISAAREFQLDRKPPIDVYSGAGEKWDNFQKIDFIDCKFTYPSRPDIQVLNGLSV 1101
 QY 1060 EVKGGOTLALVGGSGCGKSTVWOLLREFFYDPLAGSVLIDGKEIKHLNVQWLAHILGIYSQ 1119
 DB 1102 SYDPEQTLAFVGGSGCGKSTSIQLERFYDPODQGTVMIDGHSKKNVQFLRSNTGIYSQ 1161
 QY 1120 EPILFDCSIAENTAYGNSRVVSHHEEIMQAKEANIHHETLPEKYNTRVGDGKTQLSG 1179
 DB 1162 EPVLFDCSIMDNKYGDNTKEISVERAIAAQAQLHDFVMSLPKEYETNVGIGQSLSR 1221
 QY 1180 GOKORIAIARALYRPHILLDEATSDLTSEKVVQVQALDKAREGTCIVIAHRLSTIQ 1239
 DB 1222 GEXORIAIARIYRDPKILLDEATSDLTSEKVTQALDKAREGTCIVIAHRLSTIQ 1281
 QY 1240 NADLVVVFQNGKKEHGHGTHOOLLAOKGIYFSMI 1272
 DB 1282 NSDIIAVMSGVVIEKTHKMLMDQKQAYKLV 1314

RESULT 14
 ABIL_RAT
 ID ABIL_RAT STANDARD; PRT; 1321 AA.
 AC 070127;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11)
 DE (Sister of P-glycoprotein).
 GN ABCB11 OR BSEP OR SPGP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
 RX MEDLINE=98212048; PubMed=9545351;
 RA Gerloff T., Stieglar B., Hagenbuch B., Madon J., Landmann L., Roth J.,
 RA Hofmann A.F., Meier P.J.;
 RA "The sister of P-glycoprotein represents the canalicular bile salt
 RT export pump of mammalian liver."
 RL J. Biol. Chem. 273:10046-10050(1998).
 CC -!- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
 CC INTO THE CANALICULUS OF HEPATOCYTES.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
 CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
 CC MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
 CC SITU.
 CC -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
 CC ATP BINDING CASSETTE (ABC) DOMAIN.

MDRL_CABEL
 ID MDRL_CABEL STANDARD; PRT; 1321 AA.
 AC P34712;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Multidrug resistance protein 1 (p-glycoprotein A).
 GN PGP-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=93085750; PubMed=1360540;
 RA Lincke C.R., The I., van Groenigen M., Borst P.;
 RT "The p-glycoprotein gene family of *Caenorhabditis elegans*. Cloning
 RT and characterization of genomic and complementary DNA sequences.";
 RL J. Mol. Biol. 228:701-711(1992).
 RN [2]
 RP TISSUE SPECIFICITY.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=93223702; PubMed=8096815;
 RA Lincke C.R., Broeks A., The I., Plasterk H.A., Borst P.;
 RT "The expression of two p-glycoprotein (pgp) genes in transgenic
 RT *Caenorhabditis elegans* is confined to intestinal cells.";
 RL EMBO J. 12:1615-1620(1993).
 CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
 CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: INTESTINAL CELLS.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X65054; CAA46190.1; -;
 CC PIR: S27337; S27337.
 CC InterPro: IPR003593; AAA.
 CC InterPro: IPR001140; ABC_transporter_tmern.
 CC InterPro: IPR003439; ABC_transportr.
 CC InterPro: IPR001687; ATP_GTP_A.
 CC Pfam: PF00664; ABC_membrane; 2.
 CC Pfam: PF00005; ABC_tran; 2.
 CC SMART: SM00382; AAA; 2.
 CC PROSITE: PS00211; ABC_TRANSPORTER; 2.
 CC APP-binding: Glycoprotein; Transmembrane; Transport; Repeat;
 KW Multigene family.
 FT DOMAIN 1 77 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 78 98 POTENTIAL.
 FT TRANSMEM 144 164 POTENTIAL.
 FT TRANSMEM 213 233 POTENTIAL.
 FT TRANSMEM 240 260 POTENTIAL.
 FT TRANSMEM 321 341 POTENTIAL.
 FT TRANSMEM 350 370 POTENTIAL.
 FT DOMAIN 371 753 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 754 774 POTENTIAL.
 FT TRANSMEM 798 818 POTENTIAL.
 FT TRANSMEM 874 894 POTENTIAL.
 FT TRANSMEM 895 915 POTENTIAL.
 FT TRANSMEM 978 998 POTENTIAL.
 FT TRANSMEM 1017 1037 POTENTIAL.
 FT DOMAIN 1038 1321 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 451 458 ATP (POTENTIAL).
 FT NP_BIND 1112 1119 ATP (POTENTIAL).
 SS SEQUENCE 1321 AA; 145074 MW; 6130AAF3B31A8FA9 CRC64;

Query Match 42.3%; Score 2738; DB 1; Length 1321;
 Best Local Similarity 45.0%; Pred. No. 2.6e-147;
 Matches 581; Conservative 225; Mismatches 440; Indels 44; Gaps 11;
 QY 20 KSKKKEKKKPTVSTFAMFRYSNWLRLYMLVGTAAIIHGAALPLMLVFNWMTDSFA 79
 DB 47 KITRDAKEEVNVKVSIPQLYRYTTLEKLLLFGLTGLVAVITGAGLPLMSILQGVQAFI 106
 QY 80 NAGISRNKTPPVVINSITNNTHQFI-----NHLEEMTYYYYYGIGAGVLVAAY 131
 DB 107 NE-----QIVIN-----NNGSTFLPTGQNTKTDDEHDVNMVNVSYAAMTVGMAAQ 154
 QY 132 IOVSFWCLAAAGROILKIRKOFHAIMROEIGWDFDVHGVGELNTRLDVDSKINEGIDKV 191
 DB 155 IIVTCYLYVAEQMNNRLREFVKSILRQELSFDTNHSGLTATKLPDLNLERVKEGDKI 214
 QY 192 GMFQSFATFTFTGIVGTPGMKLTILVLAISPLVLSAAIWAIIKLSSTDSFTLAYAKA 251
 DB 215 GMAFYLSQFITGFIIVAFTHSQWLTIVLAVTPIQALCGFAIAKSMSTFAIRETLRYAKA 274
 QY 252 GAVAEVLAIRTVIAFGGKKELERYKNLEAKRIGIKKAITANISGAAPLLIYASY 311
 DB 275 KYVEETISIRTVSLNGLRYELERYSTAVEAKKAGVLKGLFLGIFSGMAQSNFISF 334
 QY 312 ALAFWGTSLVLSSEVTIGVLTVPFSLVIGAFSIGOASPSIEAFANARAAVEIFKIID 371
 DB 335 ALAFYIGVGVHDGSLNFGDMLTTFSSVMGSMALAGLQPLAVLGTAGGASGIYEVLD 394
 QY 372 NKPSIDYSKSGHKPDNIKGNLFKNVHFSYPSRKEVKILKGLNLKVKVSGGTVALVNSG 431
 DB 395 RKPVIDSSKAGRKDMKIGDITVENVHTYPSRPDPLRGMNLRVNAQIVALVGSG 454
 QY 432 CGKSTIVQIMQRLYDPTDGMVCDIGODIRTNVRLHREITGVVSOEPLVFATTIENRY 491
 DB 455 CGKSTIISLLRYDYDKITIDVGDVDRDINLEFLRKNVAVVSOEPALFNCTIENISL 514
 QY 492 GRENVTMDIEKAVKANAYDFIMKLPNKFDTLVGERGARLSGGQKQRTAIRALVRNPK 551
 DB 515 GREGITREEMVAACKMANNAEKIKYLPNGYNTLVGDRGTQLSGGQKQRTAIRALVRNPK 574
 QY 552 ILLDDEATSAIDSEAVVQVVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGVIYVKG 611
 DB 575 ILLDDEATSAIDSEBIVQALDKAAKGRITIIAHLRLSTIRNADLIITCKNGQWVEVG 634
 QY 612 NHDELMKEGIYFKLVMTOTRGNEIE-----LENATGESKESDSDALEMPSKDGSSL 663
 DB 635 DHRALMAQQGLYDLVTAQFTTDAVDSAAEGKFSRENSVAROTSEHGLSRQASMDIM 694
 QY 664 IK-RKSTRSI-HAPQ-----GQDRKLGTKEDLINE-NVPSVSWFRILKLNTEWPF 712
 DB 695 NVRSTIGSITNGPVIDEKEERIGKDALSRLKQLEENNAQKTNLFELLYHARPHALS 754
 QY 713 VVGIFCAIINGLQPAFSIIFSRIGITRDEDPETKQNSNMFSVFLVLGIISFIFF 772
 DB 755 FIGMSTATIGGFYPTYSYVFTSFNNVFA--GNPADFLSQGHFWALMELVLAAAQIGCSF 812
 QY 773 LOGFTFGAGETLTTRKLRVMPFRMLRODVSWDFDPKNTGTALTTRLANDAAQYKGAIGS 832
 DB 813 LMTFPMGTASESLTRDLRNKLFNRSQHIGFDFDSQNASGRISLTRLATDVPNLRTAIDF 872
 QY 833 RLAVITQNIANGTGIISLIYGWQLTLLLAIVPIIAIAGVVMKMLSGOALKDKKLE 892
 DB 873 REFTVITLVSVAGIGLAFYQWQMALIITAILPIVAFGYLRGRRTGKNVKSASEFA 932
 QY 893 GAGKATATEIENFRVSVLTREQKEFYVYQSLQVPYRNSLRKAHIFGVSSITQAMMYF 952
 DB 933 DSGKTAIEAIENVRTVQALAREDTYENFCEKLDIPHKEAFIQGLSYGCASSVLYL 992
 QY 953 SYAGCFRGAYLVANEENFQD--VLLVFAIVFCGAMAVGOVSPFADYAKAKYSAAHVI 1010
 DB 993 LNTCAVRMGLALIIDPTMQPMRVLVRVWYALITSTLTGATSTFYFETAKATAGGIIF 1052
 QY 1011 MIEKSPIDSYSPLHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEVKGQTLALV 1070

Db 1053 GMLRKISKIDSLAGEK-KKLYGKVIKFNVRFPAYPERPEIEILKGLSFSEVPGQTLAV 1111
QY 1071 GSSGCGKSTVWQLLERFYDPLAGSVLIDGKEIKHLNVQWLRHGLGIVSQEPIILFDCSIAE 1130
Db 1112 GPSGCGKSTVWALLERFYDTLGGEIFIDGSEIKTLNPEHTRSQIAIVSQEPTLFDCSIAE 1171
QY 1131 NIAYGNSRVVSHHEITMOAAKEANIIHHFIETLPEKYNTRVGDGTOLSGGQKORIAIARA 1190
Db 1172 NIYGLDPSSVTMAQVEEARLANIHNFAELPEGFETRVGDRGTOLSGGQKORIAIARA 1231
QY 1191 LVRQPHILLDEATSALDTESEKVVQEQALDKAREGRTICIVIAHRLSTIQNADLIVVFONG 1250
Db 1232 LVRNPKILLDEATSALDTESEKVVQEQALDRAREGRTICIVIAHRLNTVMNADCIAYVNSG 1291
QY 1251 KVKEHGTQOLLAQKGIYFSMISVQAGAKR 1280
Db 1292 TIIKGTHTQIMSEKGAAYKLTOKQMTERR 1321

Search completed: November 6, 2002, 18:40:22
Job time : 19.4899 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:31:39 ; Search time 29,3717 Seconds
(without alignments)
7539.009 Million cell updates/sec

Title: US-09-672-725C-4

Perfect score: 6467

Sequence: 1 MDPEGRKGSAEKNFWKMGK.....LLAQKGIYFSMISVQAGAKR 1280

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6467	100.0	1280	6	O46605	O46605 canis famil
2	5666	87.6	1285	6	O02793	O02793 ovis aries
3	5634.5	87.1	1272	11	Q9JK64	Q9JK64 rattus norv
4	5612.5	86.8	1272	11	Q924L6	Q924L6 rattus norv
5	5505.5	85.1	1163	6	Q9TSU2	Q9TSU2 felis silve
6	5321	82.3	1169	11	O60502	O60502 cricetus
7	4654	72.0	1288	13	Q93437	Q93437 gallus gall
8	4430.5	68.5	1287	13	Q91586	Q91586 xenopus lae
9	3333.5	51.5	1348	13	Q90235	Q90235 raja erinac
10	3201	49.5	1321	11	O88331	O88331 rattus norv
11	2867	44.3	1294	5	O45721	O45721 caenorhabdi
12	2821	43.6	851	13	Q9W693	Q9W693 fundulus he
13	2817	43.6	1265	5	O01495	O01495 caenorhabdi
14	2810	43.5	1275	5	O61301	O61301 haemonchus
15	2740	42.4	1321	5	Q91349	Q91349 caenorhabdi
16	2685.5	41.5	1313	5	Q9V626	Q9V626 drosophila

```
17 2636.5 40.8 1283 5 Q24393 Q24393 drosophila
18 2611 40.4 1285 10 Q9LX1 Q9LX1 oryza sativ
19 2610 40.4 1292 10 Q9M1Q9 Q9M1Q9 arabidopsis
20 2599.5 40.2 817 13 Q90WV6 Q90WV6 pseudopleur
21 2594 40.1 1278 10 Q9FWX7 Q9FWX7 arabidopsis
22 2588.5 40.0 1286 10 O80725 O80725 arabidopsis
23 2554 39.5 1279 5 Q9V616 Q9V616 drosophila
24 2533 39.2 1229 10 Q9SY12 Q9SY12 arabidopsis
25 2533 39.2 1348 3 Q9Y8G1 Q9Y8G1 emericella
26 2530 39.1 1229 10 O49749 O49749 arabidopsis
27 2520.5 39.0 1289 10 Q94IH6 Q94IH6 coptis japo
28 2519 39.0 1230 10 Q9SY13 Q9SY13 arabidopsis
29 2511.5 38.8 1349 3 O43121 O43121 aspergillus
30 2479 38.3 1229 10 Q9FWX8 Q9FWX8 arabidopsis
31 2466 38.1 1292 3 Q9C163 Q9C163 rhizomucor
32 2464.5 38.1 1313 10 Q92RG2 Q92RG2 solanum tub
33 2440 37.7 1248 10 Q9PHF1 Q9PHF1 arabidopsis
34 2434.5 37.6 1323 10 Q9M0M2 Q9M0M2 arabidopsis
35 2432.5 37.6 1408 3 O43140 O43140 cryptococcu
36 2431 37.6 1310 5 Q24852 Q24852 entamoeba h
37 2423 37.5 1286 10 Q9ZR72 Q9ZR72 arabidopsis
38 2416 37.4 1331 3 Q9HGT5 Q9HGT5 trichophyto
39 2408.5 37.2 1266 5 Q20331 Q20331 caenorhabdi
40 2404 37.2 1252 10 Q9LTX0 Q9LTX0 arabidopsis
41 2400.5 37.1 1268 5 Q20335 Q20335 caenorhabdi
42 2395 37.0 1320 5 Q9VRW2 Q9VRW2 drosophila
43 2386 36.9 1302 5 Q24851 Q24851 entamoeba h
44 2351 36.4 538 11 Q03982 Q03982 cricetus
45 2327.5 36.0 1222 10 Q9SVW6 Q9SVW6 arabidopsis
```

ALIGNMENTS

RESULT 1

```
O46605 ID O46605 PRELIMINARY; PRT; 1280 AA.
AC O46605;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE P-GLYCOPROTEIN.
GN MDR1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Fuel O., Lepage J.F., Alvinerie M., Galtier P., Pineau T.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF045016; AAC02113.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR ATP-binding; transpor.
SQ SEQUENCE 1280 AA; 141524 MW; 762DD5AFF4C73306 CRC64;
```

```
Query Match 100.0%; Score 6467; DB 6; Length 1280;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPEGRKGSAEKNFWKMGKSKKKEKKPTVSTFAMFRYSNWLDRLYMLVGTMAAIIH 60
|||||
Db 1 MDPEGRKGSAEKNFWKMGKSKKKEKKPTVSTFAMFRYSNWLDRLYMLVGTMAAIIH 60
```

QY 61 GAALPLMLVFGNMTDSFANAGISRNKTPPVIINESITNNTQHFHNLHEEMTTVAYIYS 120
Db 61 GAALPLMLVFGNMTDSFANAGISRNKTPPVIINESITNNTQHFHNLHEEMTTVAYIYS 120
QY 121 GIGAGVLVAAYIQVSPWCLAAAGROILKIRKQFFHAIMRQIEGWDFVDHVGELNTRLTDV 180
Db 121 GIGAGVLVAAYIQVSPWCLAAAGROILKIRKQFFHAIMRQIEGWDFVDHVGELNTRLTDV 180
QY 181 SKTNEGIGDKVGMFFQSIATFTFTGFTPGWKLTVLIAISPVGLSAAIWAKILSFF 240
Db 181 SKTNEGIGDKVGMFFQSIATFTFTGFTPGWKLTVLIAISPVGLSAAIWAKILSFF 240
QY 241 TDKELLAYAKAGAVAEVLAIRTVIAFGQKKELERYNKNLEEKRIKKAITANISI 300
Db 241 TDKELLAYAKAGAVAEVLAIRTVIAFGQKKELERYNKNLEEKRIKKAITANISI 300
QY 301 GAAFLLIYASALAFWYGTSLVLSSEYTTIGQVLTVPFVSVLIGAFSIGQASPSIEAFANAR 360
Db 301 GAAFLLIYASALAFWYGTSLVLSSEYTTIGQVLTVPFVSVLIGAFSIGQASPSIEAFANAR 360
QY 361 GAAEYEFKIIDNKPSIDSYKSGHKPDNKGKLEFNHVSFSPSRKEVKILKGLNLKVS 420
Db 361 GAAEYEFKIIDNKPSIDSYKSGHKPDNKGKLEFNHVSFSPSRKEVKILKGLNLKVS 420
QY 421 GQTVVALVNGSGCKSTTVQLMQRLYDPTDGMVCIQDQDITINVRHLREITGVVSOEPVL 480
Db 421 GQTVVALVNGSGCKSTTVQLMQRLYDPTDGMVCIQDQDITINVRHLREITGVVSOEPVL 480
QY 481 FATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGARLSGGQKRI 540
Db 481 FATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFDTLVGERGARLSGGQKRI 540
QY 541 AIALALVRNPKILLDDATSEAVQVVALDKARKGRTTIVIAHRLSTVRNADVIA 600
Db 541 AIALALVRNPKILLDDATSEAVQVVALDKARKGRTTIVIAHRLSTVRNADVIA 600
QY 601 GFDDGVIVERGNHDELMEKEGIYFKLVMTQTRNEIELEENATGESKSDALEMSPKDSG 660
Db 601 GFDDGVIVERGNHDELMEKEGIYFKLVMTQTRNEIELEENATGESKSDALEMSPKDSG 660
QY 661 SSLIKRSTRRSIHAPOQDKLGTEDLNENVPVSFWILKLNSTEWYFVVGIFCAI 720
Db 661 SSLIKRSTRRSIHAPOQDKLGTEDLNENVPVSFWILKLNSTEWYFVVGIFCAI 720
QY 721 INGLQPAFSIIFSRIGITRDEDPETKQNSNMFSLVFLVLIISITFTFFLQGTFFGK 780
Db 721 INGLQPAFSIIFSRIGITRDEDPETKQNSNMFSLVFLVLIISITFTFFLQGTFFGK 780
QY 781 AGEILTKRLRYMVRPSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKAIGSLAVITON 840
Db 781 AGEILTKRLRYMVRPSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKAIGSLAVITON 840
QY 841 IANLGTGIIISLIYGWOLTLALLAIVPIIAAGVEMKMLSGQALKDKKELEGAGKIATE 900
Db 841 IANLGTGIIISLIYGWOLTLALLAIVPIIAAGVEMKMLSGQALKDKKELEGAGKIATE 900
QY 901 AIENFRTVSLTREQPEYMYAQSLQVYPYRNSLRKAHIFGVFSITQAMMYFVAGCFRF 960
Db 901 AIENFRTVSLTREQPEYMYAQSLQVYPYRNSLRKAHIFGVFSITQAMMYFVAGCFRF 960
QY 961 GAYLVANEFMFQDVLVLFSAIVFAGAMAVQVSSFADYAKAKVSAAHVIMILEKSPILD 1020
Db 961 GAYLVANEFMFQDVLVLFSAIVFAGAMAVQVSSFADYAKAKVSAAHVIMILEKSPILD 1020
QY 1021 SYSPHGLKPNTELEGNTVNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKSTV 1080
Db 1021 SYSPHGLKPNTELEGNTVNEVFNYPTRPDIPVLQGLSLEVKKGQTLALVSSGCGKSTV 1080
QY 1081 VOLLERFYDPLAGSVLIDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAENIAYGDSNRV 1140
Db 1081 VOLLERFYDPLAGSVLIDGKEIKHLNVQWMLRAHLGIVSQEPILFDCSIAENIAYGDSNRV 1140
QY 1141 VSHEEIMQAQKEANIHHFETIPEKYNTRVGDGKQTLGGGQKORIAIARALVRQPHILL 1200

Db 1141 VSHEEIMQAQKEANIHHFETIPEKYNTRVGDGKQTLGGGQKORIAIARALVRQPHILL 1200
QY 1201 DEATSALDTESEKVVQBALDKAREGRTTCIVIAHRLSTIQNALDITVVPFQNGKVEHGHQ 1260
Db 1201 DEATSALDTESEKVVQBALDKAREGRTTCIVIAHRLSTIQNALDITVVPFQNGKVEHGHQ 1260
QY 1261 LLAQKGIYFMSISVOAGAKR 1280
Db 1261 LLAQKGIYFMSISVOAGAKR 1280
RESULT 2
002793
ID 002793 PRELIMINARY; PRT: 1285 AA.
AC 002793;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN-1.
GN MDRI.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Longley M., Crawford A.M.;
RT "Ovine mdrl gene."
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; U78609; AAB58489.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmam.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1285 AA; 142020 MW; 90153E617C44856F CRC64;
Query Match 87.6%; Score 5666; DB 6; Length 1285;
Best Local Similarity 86.9%; Pred. No. 0;
Matches 1117; Conservative 83; Mismatches 79; Indels 6; Gaps 3;
QY 1 MDPEGRGK-SAEKNFWKMGKK- ---SKKEKKEKPTVSTFAMFRYSNWLDRKMLVGTM 55
Db 1 MDLEGDRNGRAAGNFKRDKRRFFSKDEKKEKPTVSTFTMFRYSNWLDRCLMVLGTL 60
QY 56 AAIHGAALPLMLVFGNMTDSFANAGISRNKTPPVIINESITNNTQHFHNLHEEMTTY 115
Db 61 AAIHGAALPLMLVFGNMTDSFAGAGNLGNTLNSINTSTIDRTE-YGKKLEKEMTY 119
QY 116 AYYISGIGAGVLVAAYIQVSPWCLAAAGROILKIRKQFFHAIMRQIEGWDFVDHVGELNTR 175
Db 120 AYYISGIGAGVLVAAYIQVSPWCLAAAGROVHRIRKQFFHAIMRQIEGWDFVDHVGKLNTR 179
QY 176 LTDDVSKINEGIDGKVMFPFQSIATFTFTGFTPGWKLTVLIAISPVGLSAAIWAK 235
Db 180 LTNDVSKINEGIDGKVMFPFQSIATFTFTGFTPGWKLTVLIAIPVLGLSAAIWAK 239
QY 236 ILSFDTDKELLAYAKAGAVAEVLAIRTVIAFGQKKELERYNKNLEEKRIKKAIT 295
Db 240 ILSFDTDKELLAYAKAGAVAEVLAIRTVIAFGQKKELERYNKNLEEKRIKKAIT 299
QY 296 ANISGAAPFLIYASALAFWYGTSLVLSSEYTTIGQVLTVPFVSVLIGAFSIGQASPSIEA 355
Db 300 ANISGAAPFLIYASALAFWYGTSLVLSREYSIGQVLTVPFVSVLIGAFSIGQASPSIEA 359
QY 356 FANARGAAVEIFKIIDNKPSIDSYKSGHKPDNKGKLEFNHVSFSPSRKEVKILKGLN 415

```
Db 360 FANARGAAEYFEKIIDNKPISDSYNTGKHKPDNIKNLGNLFRNVHFHYSPRNEVAILKGLN 419
Qy 416 LKVGSGQTVALVNGSGCGKSTTVQLMQRLYDPTDGMVCIDGQDTRTINVRHLREITGVVS 475
Db 420 LKVGSGQTVALVNGSGCGKSTTVQLMQRLYDPTDGMVCIDGQDTRTINVRHLREITGVVS 479
Qy 476 QEPVLFAATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFTLVGECARLSGG 535
Db 480 QEPVLFAATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFTLVGECARLSGG 539
Qy 536 QKQRIATARALVRNPKILLDEATSDALDTESEAVVQALDKARKGRTTIVIAHRLSTVRN 595
Db 540 QKQRIATARALVRNPKILLDEATSDALDTESEAVVQALDKARKGRTTIVIAHRLSTVRN 599
Qy 596 ADVTAGDDGVIVKGNHDELMEKGYFKFLVMTQTRGNEIELENATGESKSSDALEMS 655
Db 600 ADVTAGDDGVIVKGNHDELMEKGYFKFLVMTQTRGNEIELENATGESKSSDALEMS 659
Qy 656 PKDSGSSILKRSTRRSIHAPOGQDRKLTGKEDLNENVPVSFWRIILKLNSTWEPYFVG 715
Db 660 SODSRSSILRRKSTRRSIRGSQDRKLTSTEEILDESVPVFWRIILKLNSTWEPYFVG 719
Qy 716 IFCAIINGLOPAPSIIFSRIGIIFTRDEDPETKRONSMFSLFLVLGIISITFFLQ 775
Db 720 VFCAIINGLOPAPSIIFSRIGIIFTRDNDDETNRQNSNLSLFLVLGIISITFFLQ 779
Qy 776 FTFGKAGEILTKRLRYMVFSMLRODYSWFDPKNTGTALTTRIANDAAQVKAIGSRJA 835
Db 780 FTFGKAGEILTKRLRYMVFSMLRODYSWFDPKNTGTALTTRIANDAAQVKAIGSRJA 839
Qy 836 VITQNIANLGTGIISLIYQWLTLLALLAIVPIIATAGVYEMKLSQALKDKKELEGAG 895
Db 840 VITQNIANLGTGIISLIYQWLTLLALLAIVPIIATAGVYEMKLSQALKDKKELEGAG 899
Qy 896 KIATEAIENRTVSLTREOKFMYMAQSLQVYPRNSLRKAHIFGVSFSTQAMMYFSYA 955
Db 900 KIATEAIENRTVSLTREOKFMYMAQSLQVYPRNSLRKAHIFGVSFSTQAMMYFSYA 959
Qy 956 GCFRFGAYLVANEFMFQDVLVLSAIVFCAMAVGVSSFPADYAKAKYSAAHVINIEBK 1015
Db 960 GCFRFGAYLVANEFMFQDVLVLSAIVFCAMAVGVSSFPADYAKAKYSAAHVINIEBK 1019
Qy 1016 SPLIDSYSPHGLKPNLTLEGNTVFNENYFNPTRPDIPVLQGLSLEKVGKQTLALVSSGC 1075
Db 1020 IPLIDSYSTGLKPSVTEGSAFNDVYVNPTRPDVPLRGLSLEKVGKQTLALVSSGC 1079
Qy 1076 GKSTVVOLLERFDYDPLAGSVLDGKEIKHLNVQWLRHAHGIYSQEPILFDCSTAEINAYG 1135
Db 1080 GKSTVVOLLERFDYDPLAGSVLDGKEIKHLNVQWLRHAHGIYSQEPILFDCSTAEINAYG 1139
Qy 1136 DNSRVVSHEEIMQAAKEANIHFIEPLPEKYNTRVGDGKTQSGGOKQRIATARALVROP 1195
Db 1140 DNSRVVSHEEIMQAAKEANIHFIEPLPEKYNTRVGDGKTQSGGOKQRIATARALVROP 1199
Qy 1196 HILLDEATSDALDTESEKVVQALDKAREGRTTIVIAHRLSTQIADLIIVFQNGKVEH 1255
Db 1200 HILLDEATSDALDTESEKVVQALDKAREGRTTIVIAHRLSTQIADLIIVFQNGKVEH 1259
Qy 1256 GTHQQLAOKGYFMSISVQAGAKR 1280
Db 1260 GTHQQLAOKGYFMSISVQAGAKR 1284
```

RESULT 3

Q9JK64

ID Q9JK64

AC Q9JK64; PRELIMINARY; PRT: 1272 AA.

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE MULTIDRUG RESISTANCE PROTEIN 1A.

GN PGY1.

```
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=LIVER;
RA Hooiveld G.J.E.J., Wilms J.W.J., Hagenbuch B., Jansen P.L.M.,
RA Meijer D.K.F., Muller M.;
RT "Cloning and functional characterization of the rat multidrug
RT resistance protein Mdr1a.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF257746; AAF69007.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC transporter_tmem.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1272 AA; 140326 MW; B5C1DF89DFC986C5 CRC64;
```

Query Match 87.1%; Score 5634.5; DB 11; Length 1272;
Best Local Similarity 86.3%; Pred. No. 2.8e-318;
Matches 1105; Conservative 82; Mismatches 84; Indels 9; Gaps 1;

```
Qy 1 MDPEGSRKSAKSNKWKMGKSKKSKKKKPGVSVFAMFRYSNWLDRYMLVGTWAAIIH 60
Db 1 MELEDNLGRADNFSKMGKSKKKKPAVSVLTFRYAGWLDRLYMLLGTAAIIH 60
Qy 61 GAALPLMLLVFGNDTDFANAGISRNKTFPPIINESITNTQHFHNLHEEMTTVAYYS 120
Db 61 GIALPLMLLVFGNDTDFANVGNNSMSF-----YNATDIYAKLEDEMTVAYYT 111
Qy 121 GTGAGVLYAAYIQVSWFCLAGRLKIRKOPFFHAIMQOEIGWFDVHDVGEINRLTDDV 180
Db 112 GTGAGVLYAAYIQVSWFCLAGRLKIRKOPFFHAIMQOEIGWFDVHDVGEINRLTDDV 171
Qy 181 SKINIGIGDKVGMFQSTATFTFGTGVKLTVLILALSPVLGSLAAWAKILSSF 240
Db 172 SKINIGIGDKVGMFQSTATFTFGTGVKLTVLILALSPVLGSLAAWAKILSSF 231
Qy 241 TKELLAYAKAGAAVEELAAIRTVIAFGQKKELERYNKLEAKRIGIKKAITANISI 300
Db 232 TKELLAYAKAGAAVEELAAIRTVIAFGQKKELERYNKLEAKRIGIKKAITANISM 291
Qy 301 GAAFLIIYASVALAFWYGTSLVSEYTIQVLTAVFVSVLIGAFSGQASPIAFANAR 360
Db 292 GAAFLIIYASVALAFWYGTSLVSEYTIQVLTAVFVSVLIGAFSGQASPIAFANAR 351
Qy 361 GAAVEFIKIDNKPISDSYSGSHKPDNIKNLFGNHFVSPSRKEVKILKGLNKLAVOS 420
Db 352 GAAVEFIKIDNKPISDSYSGSHKPDNIKNLFGNHFVSPSRKEVKILKGLNKLAVOS 411
Qy 421 GQTVLVGNSCGKSTTVQLMQRLYDPTDGMVCIDGQDTRTINVRHLREITGVVSQEPVL 480
Db 412 GQTVLVGNSCGKSTTVQLMQRLYDPTDGMVCIDGQDTRTINVRHLREITGVVSQEPVL 471
Qy 481 FATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFTLVGECARLSGGQKQRI 540
Db 472 FATTIAENIRYGRNVTMDIEKAVKEANAYDFIMKLPNKFTLVGECARLSGGQKQRI 531
Qy 541 AIARALVRNPKILLDEATSDALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVA 600
Db 532 AIARALVRNPKILLDEATSDALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVA 591
Qy 601 GFDGCVIVEKGNHDELMEKGYFKFLVMTQTRGNEIELENATGESKSSDALEMSPKDS 660
Db 592 GFDGCVIVEKGNHDELMEKGYFKFLVMTQTRGNEIELENATGESKSSDALEMSKDSG 651
```



```
QY 1141 VSHEEIMOAKAEANHFIETLPKYNTRVGDGKGTQLSGGOKQRIATARALVROPHILL 1200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1132 VSHEEIVAAKEANHFIETLPKYNTRVGRDKGTQLSGGOKQRIATARALVROPHILL 1191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1201 DEATSALDTESEKVVQEAALDKAREGTCIVIAHRLSTIONADLIWVFGNGVKKEGTHQO 1260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1192 DEATSALDTESEKVVQEAALDKAREGTCIVIAHRLSTIONADLIWVFGNGVKKEGTHQO 1251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 1261 LLAQGIYFMSISVQAGAKR 1280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1252 LLAQGIYFMSVQAGAKR 1271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 5
Q9TSU2
ID Q9TSU2 PRELIMINARY; PRT; 1163 AA.
AC Q9TSU2;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE MULTI-DRUG RESISTANCE RELATED (FRAGMENT).
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
SEQUENCE FROM N.A.
RA Okai Y., Nakamura N., Matsusiro H., Kato H., Setoguchi A., Yazawa M.,
RA Okuda M., Watarai T., Hasegawa A., Tsujimoto H.;
RT "Molecular analysis of multidrug resistance in lymphoma cells in the
RT cat.";
RL Am. J. Vet. Res. 0:0-0(1999).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AB029153; BRA87071.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
FT NON_TER 1
FT TER 1163
SQ SEQUENCE 1163 AA; 128511 MW; 1B5B413776A93A26 CRC64;

Query Match 85.1%; Score 5505.5; DB 6; Length 1163;
Best Local Similarity 93.9%; Pred. No. 6.5e-311;
Matches 1093; Conservative 37; Mismatches 33; Indels 1; Gaps 1;

QY 60 HGAALPLMLVFGNMTDSFANAGISRNKTPPVLIINESITNTQHFTHNLEEMTTVAYY 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 HGAALPLMLVFGNMTDSFANAGISRNLTTLNTGESVINDS-YFINRLEEEMTTVAYY 59
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 120 SGTCAGVLVAAYIQVFWCLAAGRQILKIRKQFFHAIMRQEIGWFDVHDVGNLRLTDD 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 60 SGTCAGVLVAAYIQVFWCLAAGRQILKIRKQFFHAIMRQEIGWFDVHDVGNLRLTDD 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 180 VSKINIGIDKVGCMFQSTATPTFTGVTGKTLVILAIISPVLGISAIAWAKILSS 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 120 VSKINIGIDKVGCMFQSTATPTFTGVTGKTLVILAIISPVLGISAIAWAKILSS 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 240 FTDKELLAYAKAGAAVEELAAITVAFGGQKKEERYNKNLEAKRIGIKKAITANIS 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 180 FTDKELLAYAKAGAAVEELAAITVAFGGQKKEERYNKNLEAKRIGIKKAITANIS 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 300 IGAFLLIYASALAFWYGTSLVLSSEYTTGQVLTVFFSVLIGAFSIGQASPSIEAFANA 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 240 IGAFLLIYASALAFWYGTSLVLSSEYTTGQVLTVFFSVLIGAFSIGQASPSIEAFANA 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 360 RGAAYEIFKIIDNKPSIDSYSGHKNPDNKGLEFKNVHFSPSRKEVKILKGLNLKQV 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```

OX NCBI_TaxID=10031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA MEDLINE=91154265; PubMed=1671863;
RA Devine S.E., Hussain A., Davide J.P., Melera P.W.;
RT "Full length and alternatively spliced ppp-1 transcripts in multidrug
RT resistant Chinese hamster lung cells.";
RL J. Biol. Chem. 266:4545-4555(1991).
CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: M59254; AAA37005.1; -.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC transporter_tmern.
DR InterPro: IPR003439; ABC transportr.
DR InterPro: IPR001687; ATP_GRP_A.
DR Pfam: PF00664; ABC_membrane; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1169 AA; 128939 MW; 72E25B7CE29DC185 CRC64;

Query Match      82.3%; Score 5321; DB 11; Length 1169;
Best Local Similarity 88.9%; Pred. No. 3.3e-300;
Matches 1039; Conservative 73; Mismatches 57; Indels 0; Gaps 0;

QY 112 MTTYYAYYSIGAGLVAAAYIQVSFWCLAAAGROIILKIRKOFFHAIMROEIGWFDVHDVGE 171
DB 112 MTTYYAYYTIGAGLVIAVYIQVSFWCLAAAGROIHKIRKOFFHAIMNQEIGWFDVHDVGE 60

QY 172 LNTRLTDDVSKINEGIDKGVMEFFQSFIATPFTTIGVFGTTPGKWLTLVLAISPVLSAA 231
DB 172 LNTRLTDDVSKINEGIDKGVMEFFQAWYFFGGFIIGFTGKWLTLVLAISPVLSAG 120

QY 232 IWAKILSSFTDKELLAYAKAGAAVEVLAIRTVIAFGGOKKELRYKNKLNLEAKRIGIK 291
DB 121 IWAKILSSFTDKELQAYAKAGAAVEVLAIRTVIAFGGOKKELRYNNLEAKRLGIK 180

QY 292 KATTANISGAFFLIIVASVALAFWGTSLVLSSEYTTIGQVLTVFFSVLIGAFSIGQASP 351
DB 181 KATTANISGAFFLIIVASVALAFWGTSLVLSKEYSIGQVLTVFFAVLAFPIGQASP 240

QY 352 SIEAFANAGAAVEIFKIIDNKPSIDSYSGSHKPNKIGNLEFNKFNHFSYPSRKEVKIL 411
DB 241 NIEAFANAGAAVEIFNIIDNKPSIDSFSGKNGYKPNKIGNLEFNKFNHFSYPSRQDVQIL 300

QY 412 KGLNLKVSQGTVALVNSCGGKSTTVOLMORLYDPTDGMVCDIGQDIRTINVRHLREIT 471
DB 301 KGLNLKVSQGTVALVNSCGGKSTTVOLLQRLYDPTGEGVVSIDGQDIRTINVRHLREII 360

QY 472 GVVSQEPVLPATTIAENIRYGRNVMTDETEKAVKEANAYDFIMKLPNKFDTLVGERGAR 531
DB 361 GVVSQEPVLPATTIAENIRYGRNVMTDETEKAVKEANAYDFIMKLPNKFDTLVGERGAQ 420

QY 532 LSGGOKQRIARALVRNPKILLDEATSDALDTESEAVVOALDKARKGRTTVIAHRLS 591
DB 421 LSGGOKQRIARALVRNPKILLDEATSDALDTESEAVVOALDKAREGRTTVIAHRLS 480

QY 592 TVRNADVIAGFDGCVIVEKGNHDLMEKGIYFKPLVTMOTRGNEIELENATGESKESDA 651
DB 481 TVRNADIAGFDGCVIVEQGNHEELMEKGIYFKPLVMTQTAGNEIELENGEVGSKNEIDN 540

QY 652 LEMSPKDGSSGLIKRRSTRIRSIHAPQGDQRKLGTKEDLNENVPVSFWIRILKLNSTWPY 711
DB 541 LDMSSKDSASSLIRRRSTRIRSIHAPQDRKLSLTKALDEDDVPPISFWIRILKLNSTWEPY 600

QY 712 FVVGIFCAITNGGLQAPAFSIFSRIGIFRDEDEPEKRONSNWFSVLFLVLGLISFIF 771
DB 601 FVVGIFCAITNGGLQAPAFSIFSRVGVFTRNTDDETEKRDHSDNLSLFLVLGLISFIF 660

QY 772 FLOGFTFGKAGEIITKLRYNVFMSRLRQDVSFDPDKNTTGALTTRLANDAAQVGAIG 831
DB 661 FLOGFTFGKAGEIITKLRYNVFMSRLRQDVSFDPDKNTTGALTTRLANDAQVKGATG 720
```

```

QY 832 SRLAVITQNIANLGTGIIISLIYGWQLTLLLLAIVIPPIAIAGVVMKMLSGQALKKKEL 891
DB 721 ARLAVITQNIANLGTGIIISLIYGWQLTLLLLAIVIPPIAIAGVVMKMLSGQALKKKEL 780

QY 892 EGAGKIATEAIENPRTVVSLTREQKFYMYAQSLQVYPYRSLRKAHFHFGVSFSTQAMMY 951
DB 781 EGSKGIATEAIENPRTVVSLTREQKFENYMQSLQIPYRNALKKAHVFGIIFSTQAMMY 840

QY 952 FSYAGCFRFGAYLVANEFMNFQDVLVFSALVFGAMAVGVSSFPADYAKAKVSAAHVIM 1011
DB 841 FSYAACFRFGAYLVARELMTFENVLLVFSALVFGAMAVGVSSFPADYAKAKVSASHIM 900

QY 1012 IIEKSPILDSYSPHGLKPNLTLEGNVNEVFNTPRPDPVLOGLSLEFKVKGOTLALVG 1071
DB 901 IIEKVPSIDSYSTGGLKPNLTLEGNVNEVFNTPRPDPVLOGLNLEVKKGOTLALVG 960

QY 1072 SSGCGKSTVVOLLERFVDPPLAGSVLDGKEIKHLNQLVWLAHGLGIVSQEPLFPCSTAE 1131
DB 961 SSGCGKSTVVOLLERFVDPPLAGSVLDGKEIKHLNQLVWLAHGLGIVSQEPLFPCSTAE 1020

QY 1132 IAYGNSRVVSHHEIMOAAKEANIHHFTETLPEKYNTRVGDGKTQLSGGOKQRIAIARAL 1191
DB 1021 IAYGNSRVVSDLETERAKKANIHQFIESLPDKYNTRVGDGKTQLSGGOKQRIAIARAL 1080

QY 1192 VRQPHILLDEATSDALDTESEKVVQEQALDKAREGTCIVIAHRLSTIONADLIYVFQNGK 1251
DB 1081 VRQPHILLDEATSDALDTESEKVVQEQALDKAREGTCIVIAHRLSTIONADLIYVQNGK 1140

QY 1252 VKEHGTHOOLLAOKGIYFSMISVQAGAKR 1280
DB 1141 VKEHGTHOOLLAOKGIYFSMISVQAGAKR 1169

RESULT 7
O93437 PRELIMINARY; PRT; 1288 AA.
AC O93437;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ABC TRANSPORTER PROTEIN.
GN CMORL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RX MEDLINE=99209805; PubMed=10195430;
RA Edelmann H.M.L., Duchek P., Rosenthal F.E., Foeger N., Glackin C.,
RA Kane S.E., Kuchler K.;
RT "Cmrl, a chicken P-glycoprotein, confers multidrug resistance and
RT interacts with Estradiol";
RL Biol. Chem. 380:231-241(1999).
DR EMBL: AJ009799; CAA08835.1; -.
DR HSP; P13569; INBD.
FT CHAIN 2 1288 ABC TRANSPORTER PROTEIN.
SQ SEQUENCE 1288 AA; 141917 MW; CB258A5F2826DB6C CRC64;

Query Match      72.0%; Score 4654; DB 13; Length 1288;
Best Local Similarity 71.0%; Pred. No. 1.9e-261;
Matches 916; Conservative 164; Mismatches 166; Indels 44; Gaps 12;

QY 2 DPEGGRKGSAAENFWMKSKKSKKKEKKP-TVSTFAMFRYSNWLDRMLYMLVGTMAAIIH 60
DB 27 DPEDEKKG-----KSKKSKKSKKSKKSKKSKKSKKSKKSKKSKKSKKSKKSKKSKKSKK 72

QY 61 GAALPLMLVFGNMVTDSPANAGISRNKFPFVILNESINNTQHTFNLHLEEMTTAYYS 120
DB 73 GTSPLAMIIFGMDTDSFVTSMTN-----ITGNSGLNSADVFNKLEEMTRAYYS 127
```

QY 121 GIGAGLVAAAYIQVSFWCLAAAGROILKIRKOFFHAIMROELIGWDFVDHVGELNTRLTDDV 180
 DB 128 AIAAAVVAAYIQVSFWCLAAAGROILKIRKOFFHAIMROELIGWDFVDHVGELNTRLTDDV 187
 QY 181 SKINIGIGDKVGMFFQSIATFTFTGIVGTPGWKLTIVILAIISPVGLSAAIAWAKILSSP 240
 DB 188 SKINIGIGDKVGMFFQSIATFTFTGIVGTPGWKLTIVILAIISPVGLSAAIAWAKILSSP 247
 QY 241 TDKELLAYAKAGAAVEEVLAIIRVIAFGGOKKELERYKNLEAKRIGIKKAITANISI 300
 DB 248 TDKEAAVAKAGAAVEEVLAIIRVIAFGGOKKELERYKNLEAKRIGIKKAITANISI 307
 QY 301 GAAFLIIYASALAFWYGTSLVLSSEYTIQOVLTVPFVSLVIGAFSGOASPSIEAFANAR 360
 DB 308 GAAFLIIYASALAFWYGTSLVLSSEYTIQOVLTVPFVSLVIGAFSGOASPSIEAFANAR 367
 QY 361 GAAVEIFKIDNKFSIDSYSGSKHGPDKNIRKGNLEFKNVHFSYPSRKEVKILKLNKVQS 420
 DB 368 GAAVAIFNIIDNEPIDSYSDAGHKPHIKGNLEFQNVFNPSRDPDVEILKLNKVNC 427
 QY 421 GQTVLVGNSCGKSTTVQLMORLYDPTDGMVCIQDQDITINVRHLREITGVVSOEPVL 480
 DB 428 GQTVLVGNSCGKSTTVQLMORLYDPTDGMVCIQDQDITINVRHLREITGVVSOEPVL 487
 QY 481 FATTIAENIRYGRNVMTDEIEKAVKANAAYDFIMKLPNKFDTLVGEGARLSGGQKORI 540
 DB 488 FATTIAENIRYGRNVMTDEIEKAVKANAAYDFIMKLPNKFDTLVGEGARLSGGQKORI 547
 QY 541 AIAALVYRNPKILLIDATSAIDTESAVVQVALDKARKGRTTIVIAHRLSTVRNADVIA 600
 DB 548 AIAALVYRNPKILLIDATSAIDTESAVVQVALDKARKGRTTIVIAHRLSTVRNADVIA 607
 QY 601 GFDDGVIVKGNHDELKKEGIYFKLVMTQTRGNELEENATCESKESDALEMPSPKSG 660
 DB 608 VFESGVITEGNSHSLIEKKGIYFKLVMTQTRGNELEENATCESKESDALEMPSPKSG 659
 QY 661 S-----SL-----TKRSTRSIH---APOQDQKLGTKEDLNENVPVSFWIRILKINSTE 708
 DB 660 SQSNLDESKELRRGSTRSMKKPGEPNDTDEK-GSSPD--EELPPVSFLKMLKLNKE 716
 QY 709 WPFVVGIVGICAITINGLOPAFSIFRSRIIGITREDDPTKQNSMPSVLEVLVIGIISF 768
 DB 717 WPFVVGIVGICAITINGLOPAFSIFRSRIIGITREDDPTKQNSMPSVLEVLVIGIISF 775
 QY 769 ITFFLGFTFGKAGEILTKRLRYWPRSMRLQDVSWFDDPKNTGTALTRLANDAAQVK 828
 DB 776 ITFFLGFTFGKAGEILTKRLRYWPRSMRLQDVSWFDDPKNTGTALTRLANDAAQVK 835
 QY 829 AIGSLAVITONANTIGTIIISLYGWOLTLALLAIVPIIAIAGVEMKMLSGOALKDK 888
 DB 836 AIGSLAVITONANTIGTIIISLYGWOLTLALLAIVPIIAIAGVEMKMLSGOALKDK 895
 QY 889 KELEGAGKATEAENIRVWISLTREKPEYMAQSLOVPYNSLRKAHIFGVFSITQA 948
 DB 896 KELEGAGKATEAENIRVWISLTREKPEYMAQSLOVPYNSLRKAHIFGVFSITQA 955
 QY 949 MMTFSYAGCFRFGAYLVNAGNEFMNFQDLVLSAIVFAGAMVQVSSFPADYAKAKVSAH 1008
 DB 956 MMTFSYAGCFRFGAYLVNAGNEFMNFQDLVLSAIVFAGAMVQVSSFPADYAKAKVSAH 1015
 QY 1009 VIMIIEKSLIDSYSPHGLKPNLTLEGNTVFNVEVNYTRDPDIPVLQGLSLEVKGOTLA 1068
 DB 1016 VIMIIEKSLIDSYSPHGLKPNLTLEGNTVFNVEVNYTRDPDIPVLQGLSLEVKGOTLA 1075
 QY 1069 LVGSSCGKSTVQLLEFYDPLAGSLVDGKEIKHLNVOLRAHLGTVVSOEPLFPCSI 1128
 DB 1076 LVGSSCGKSTVQLLEFYDPLAGSLVDGKEIKHLNVOLRAHLGTVVSOEPLFPCSI 1135
 QY 1129 AENIAYGNSRVVSHIEIQAKEANIHHTIETPEKYNTRVGDGTQLSGGQKORITAI 1188
 DB 1136 AENIAYGNSRVVSHIEIQAKEANIHHTIETPEKYNTRVGDGTQLSGGQKORITAI 1195

QY 1189 RALVROPHILLDEATSAIDTESKVVQVQALDKAREGRTCTIVIAHRLSTIONADLIWVQ 1248
 DB 1196 RALVROPHILLDEATSAIDTESKVVQVQALDKAREGRTCTIVIAHRLSTIONADLIWVQ 1255
 QY 1249 NGKVEHCHTHOOLLAQKGIYFMSLSVQAGA 1278
 DB 1256 NGKVEHCHTHOOLLAQKGIYFMSLSVQAGA 1285
 RESULT 8
 Q91586 PRELIMINARY; PRT; 1287 AA.
 ID Q91586
 AC Q91586
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE MULTIDRUG RESISTANCE PROTEIN.
 GN XEMDR.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95322451; PubMed=7599185;
 RA Castillo G., Shen H.J., Horwitz S.B.;
 RT "A homologue of the mammalian multidrug resistance gene (mdr) is
 RT functionally expressed in the intestine of Xenopus laevis.";
 RL Biochim. Biophys. Acta 1262:113-123(1995).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; U17608; AAA75000.1; -;
 DR InterPro; IPR003593; AAA.
 DR InterPro; IPR001140; ABC_transporter_tmem.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR001687; ATP_GTP_A.
 DR Pfam; PF00664; ABC_membrane_2.
 DR Pfam; PF00005; ABC_tran; 2.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
 KW ATP-binding; Transport.
 SQ SEQUENCE 1287 AA; 141505 MW; 06E95073C5771415 CRC64;
 Query Match 68.5%; Score 4430.5; DB 13; Length 1287;
 Best Local Similarity 67.6%; Pred. No. 1.8e-248;
 Matches 870; Conservative 174; Mismatches 210; Indels 33; Gaps 11;
 QY 2 DPEGGRKSAEKNFWKMGKSKKKEKKPTVSTFAMFYSNWLDRMLVMTAAIITHG 61
 DB 21 DPNSNSK--EKKGFSSKFK--KKEKTEKPKVGVTFMFRYSSTSKMLMFGTASLAHG 77
 QY 62 AALPLMLVFGNMTDSFANAG--ISRNKTFPVIINESNTNTQHFINHEEMTYAYY 119
 DB 78 AALPLMLVFGNMTDSFANAG--ISRNKTFPVIINESNTNTQHFINHEEMTYAYY 128
 QY 120 SGICAGLVAAAYIQVSFWCLAAAGROILKIRKOFFHAIMROELIGWDFVDHVGELNTRLTDD 179
 DB 129 SGICAGLVAAAYIQVSFWCLAAAGROILKIRKOFFHAIMROELIGWDFVDHVGELNTRLTDD 188
 QY 180 VSKINIGIGDKVGMFFQSIATFTFTGIVGTPGWKLTIVILAIISPVGLSAAIAWAKILSS 239
 DB 189 VSKINIGIGDKVGMFFQSIATFTFTGIVGTPGWKLTIVILAIISPVGLSAAIAWAKILSS 248
 QY 240 FTDKELLAYAKAGAAVEEVLAIIRVIAFGGOKKELERYKNLEAKRIGIKKAITANIS 299
 DB 249 FTDKELLAYAKAGAAVEEVLAIIRVIAFGGOKKELERYKNLEAKRIGIKKAITANIS 308
 QY 300 IGAFLIIYASALAFWYGTSLVLSSEYTIQOVLTVPFVSLVIGAFSGOASPSIEAFANAR 359
 DB 309 IGAFLIIYASALAFWYGTSLVLSSEYTIQOVLTVPFVSLVIGAFSGOASPSIEAFANAR 368
 QY 360 RGAAYEIKFIIDNKFSIDSYSGSKHGPDKNIRKGNLEFKNVHFSYPSRKEVKILKLNKVQ 419

Db 369 RGAAYTFIINQPKIDSFSEKGLKPKDKIKGDIKFEKNVIFTYPSRKDIQVLKGLNLNP 428
Qy 420 SGOVALVNSGCGKSTTVQLMORLYDPTDGMVCIDGQDIRTINVRHLEITGVVSQBPV 479
Db 429 SCKTVALVSSGCGKSTTVQLIQRFYDPEDGVITLDGQDIRSLNIRYLREITGVVSQBP 488
Qy 480 LFATTAENIRYGRNWTWDETEKAVKEANAYDFIMKLPNKFDTLVGERGARGLSGQKOR 539
Db 489 LFTTITADNIRYGRNWTWDETEKAVKEANAYDFIMKLPNKFDTLVGERGARGLSGQKOR 548
Qy 540 IATARLVNPKTLILDEATSAIDTESEAVVVALDKARKGRTTIVIAHRLSTVRNADVI 599
Db 549 IATARLVNPKTLILDEATSAIDTESEAVVVALDKARKGRTTIVIAHRLSTVRNADVI 608
Qy 600 AGFDGIVVEKGNHDELMKEKGIYFLVMTQ-----RGNEIELENATGSKSESDALEMS 655
Db 609 AGFDNGVIVGQSHKELMERGGVYFNLVLTQVETSKDEEDLETHIYEKK-----IP 661
Qy 656 PKDGSGLIKRRSTRSIHA--PQODRKLGTKE-DLNNVPSVSPWRILKLNSTWEPYF 712
Db 662 VTHSHNLVRKSRNTIKSVPEDEKEDVEDEKKKEGPPVPVFFKVKMLNKPWPYF 721
Qy 713 VVGFCIAINGLQPAFSIIFRIIGITFTRDEDPETK-RQNSMFSVLFLVLIISFITF 771
Db 722 VVGVICAMINGATQAPAFIIFSRIGVFA--GPVSQMRSESSMYSLLEFALGGSFIF 778
Qy 772 FLOGFTFGKAGEILTKRLRNVWFRSMRLQDVSWFDDPKNTTGALTTRLANDAAQVKGATG 831
Db 779 FLOGFTFGKAGEILTKRLRNVWFRSMRLQDVSWFDDPKNTTGALTTRLANDAAQVKGATG 838
Qy 832 SRLAVITONTANTGTGIIISLIYGWOLTLALLAIPTIAIAGVVEKMLSGQALKDKKEL 891
Db 839 TRALLAQNVANLGTALIIISFIYGWOLTLALLAIPTIAIAGVVEKMLSGQALKDKKEL 898
Qy 892 EGAKIATEAENPRTVVSUTREKQFYMYAQSLQVPYRNSLKAHIFGVSPITOAMMY 951
Db 899 EKAGKISTDAVLRNTRVSLTRKFEAMYEKSLGEPYRNSIKKAHLGLTYGLSQAHHV 958
Qy 952 PSYACFR-FGAYLVANFNFOVDLVAISAIVFGAMVGVSSFPADYAKAKVSAHVI 1010
Db 959 LCLCWFSVLGAYLVVGLMKLDEVLVSSAIVLGMALGQTSFPADYTKAMISAHIF 1018
Qy 1011 MIEKSPILDISYSPHGLKPNLTLEGNVTFNEVFNYPTRPDIPVLQGLSLEKVKGQTLALV 1070
Db 1019 SLLERVPQIDSYPDGEKPKNCNCGNVFKGVFNYPTRPDITVLQGLDISVKQGETIALV 1078
Qy 1071 GSSCGGKSTTVQLLERFYDPLAGSVLDGKEIKHLNVLQWLAHGLIVSQEPILFDCSIAE 1130
Db 1079 GSSCGGKSTTVQLLERFYDPLAGSVLDGKEIKHLNVLQWLAHGLIVSQEPILFDCSIAE 1138
Qy 1131 NIAYGDSRVVSHHEIMOAAKEANIHFIEITLPEKYNTRVGRDKGTQLSGGQKQRIATARA 1190
Db 1139 NIAYGDSRVVSHHEIMOAAKEANIHFIEITLPEKYNTRVGRDKGTQLSGGQKQRIATARA 1198
Qy 1191 LVROPHILLDEATSAIDTESEKVVQBALDKAREGRTTCIVIAHRLSTIQNADLIVFQNG 1250
Db 1199 LIRKPKILLDEATSAIDTESEKVVQBALDKAREGRTTCIVIAHRLSTIQNADLIVFQNG 1258
Qy 1251 KVKEGHTHQALLQKGIYFSMISVQAG 1277
Db 1259 KVVEGHTHQALLQKGIYFSMISVQAG 1285
RESULT 9
Q90235 PRELIMINARY; PRT; 1348 AA.
AC Q90235
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BILE SALT EXPORT PUMP.
OS Raja erinacea (Little skate).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;

OC Elasmobranchii; Squala; Hypnosqualea; Pristiorajae; Batoidea;
OC Rajiformes; Rajidae; Raja.
OX NCBI_TaxID=7782;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21340059; PubMed=11447010;
RA Cai S.Y., Wang L., Ballatori N., Boyer J.L.;
RA "Bile salt export pump is highly conserved during vertebrate evolution
RT and its expression is inhibited by PFIC type II mutations.";
RL Am. J. Physiol. Gastrointest. Liver Physiol. 281.G316-G322(2001).
DR EMBL; AF367243; AAK52958.1; -
SQ SEQUENCE 1348 AA; 148648 MW; 81906F641654CFDD CRC64;
Query Match 51.5%; Score 3333.5; DB 13; Length 1348;
Best Local Similarity 49.9%; Pred. No. 8.8e-185;
Matches 662; Conservative 250; Mismatches 356; Indels 59; Gaps 9;
Qy 1 MDPEGRKSGAEKFNWKMKGKSKKE---KKEKKPTVTFAMFRYSNWLDRILMLVGTMAA 57
Db 27 IDGPGQRKVTVNGNYPKSKDDNENLSKGDRIKIGFFQLFREAFASYEVTLMVFGVCA 86
Qy 58 IIHGAALPLMMLVFCNMTDSPANAGISRNTFPVILINESITNNTQHFIN----- 106
Db 87 LLHGAQPAVLLVFLGLTDTFIADYIELOELKDT--RKCDNTHIHWNGSEYLNQHNET 144
Qy 107 -----HLEEMTTVAYYYSIGAGVLVAAYIQVSWFCLAAAGROILKIRKOFFHAIMRQE 160
Db 145 MSCGILNTEKEMTFAYYYVVGICSVLLGYFQICFWYTAARQTKIRKAYFQIMRME 204
Qy 161 IGWFDVHDVGEINLTLDVSKINSGIGDKVGMFPQSFATFTFGTIVGFTPGWKILTVIL 220
Db 205 MGWFDNVSVELNTRMSDDINKINDAIADQVIGFIORETFVSGFLMGFVNGKLTVII 264
Qy 221 ATSPVLGSAALIAKILSFTDKELLAYAKAGAAVEVLAARIVTAPGOKKELERYNK 280
Db 265 AVSPLIGLGAALMALSVARLTGLGLKAYAKAGAVADEVLSSTRTVAAFSEKKEVDYDR 324
Qy 281 NLEEKRGIGIKAITANISIGAAFLLIYASALAPWYGTSLV-SSEYTIQGVLTVPFSV 339
Db 325 NLVFAQRWIRKMGITMGFTGYVWMIIFCCYALAFWYSGKLVIEQNETPGGLLQVFLGV 384
Qy 340 LGAGSIGQASPSIEAFANARAAAYEIFKIIDNKPISIDSYSGKHGKPNIKGNLEFKNVH 399
Db 385 LVAAAMLGQASCLAEAFASGRGAAYKIFETIDREPIDCMSEGGYTLNKKVGDIEFHNVT 444
Qy 400 FSYPSKEVKILKGLNLKLVQSGQTVALVNSCGGKSTTVQLMORLYDPTDGMVCIDGODI 459
Db 445 FNVPSPDVKTLDRLSMVIKAGETTAFVPGSGSGKSTAVELIQRFYDPKQGMVTLJDGHI 504
Qy 460 RINVRHLREITGVYSOEPLVFAITAEINIRYGRNVTMDTEKAVKEANAYDFIMKLPN 519
Db 505 RSLNIQWLSLIGIVEPEPLVFTTEENIRVVRGLKNDIVQAAKEANAYDFIMDLPE 564
Qy 520 KFDTLVGERGARGLSGQKQRIATARALVRNPKILLDEATSAIDTESEAVVVALDKARK 579
Db 565 KFTLVGEGGQSGGQKQRIATARALVRNPKILLDMATSAIDNESEAVIAGALDKVRF 624
Qy 580 GRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMKEKGIYFLVMTQTRGNEIELE 639
Db 625 GRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMKEKGIYFLVMTQTRGNEIELE 684
Qy 640 NATGESKESDALEMSPKDGSGLIKRRSTRSIHAPOQGDQRKLGTKEDLNE----- 691
Db 685 KA-----ROVNGALEDGASEK-RQLIRGSSRASVSTRHRSRSQVSEVLSLDSGPGDVA 739
Qy 692 ---NVPSVSFW-----RIILKNSTWEPYFVGVIFCAILINGLOPAF 729
Db 740 SAVRTPSISLGDGDKDVEEESIEBPAPVSRILKNVSEWPMYLMFGSLGAANGVGNPIY 799
Qy 730 SIIFSRLIGFTFREDPEPKRQNSMFSVLFLVLIISFITFPGFTFGKAGEILTKRL 789
Db 800 ALLFSQILGTSL-QNEEEKINQINAIICLFFVVVGLVSLTQFLQSYFFAKSGELLTRRL 858

```
QY 790 RYMFRSMLRDVSWFDDPKNTGALTTRLANDAAOVKGAIGSLAVITONIANLGTGII 849
Db 859 RKLGFQOAMLRQEIWFDDKNSPGLTTRLATDASOVGATGTQIGMTVNSITNIGVSLI 918
QY 850 ISLYGWQITLULLAIPIIAGVVEKMLSGQALDKKLEGAGKTIATIAENFRIVV 909
Db 919 TAFYFSWKLTVILCFPLPALTGAQARMLTGFANQDKEALEAAGQISSEALSNIRTIA 978
QY 910 SITRQKFEYMYAOSLOVPYRNSLRKAHIFGVFSFISITQAMMYFSGCFRGAYLVANEF 969
Db 979 GLAKEKMEVQLEAEQLEAPYKAAIKANVYGLCFGFAQSVIFLANAASYRFGGYLVVAEN 1038
QY 970 MNFQDVLVSAIVFGAMAVGVSSFPADYAKAKVSAHVIMIEKSPIDSYPHGLKP 1029
Db 1039 LHFSTVFRVSAIVTSGTALGRASSYTPDYAKAKIAAARFPQLIDRAPKINIDSEGEKW 1098
QY 1030 NTLGNVTFNEVNPYRDPDIPVLOGLSLEVKGQOTIALVSGSGCGKSTVVQLLERYD 1089
Db 1099 SNFRGDIKFVCTFTYTPRPDFQVNLGSLVCVMPGOTLAFVSGSGCGKSTVQLLERYD 1158
QY 1090 PLAGSVLIDGKEIKHLNVQWLRHLGIVSQBPILFDGSIABENIAYGNSRVVSHETMQA 1149
Db 1159 PNQGRVIDGSTVNTWNTAFRLSKIGIVSQBPVLFDSIEENIKYDONSQVSMDEVIAA 1218
QY 1150 AKENIHIFETLPEKYNTRVGDGTQOLSGGQKORIAIARALVRQPHILLIDDEATSLDT 1209
Db 1219 AKKAQLHSFVMSLPEKYGTRVGAQGSQSLRGQKORIAIARALVRDPKILLIDDEATSLDT 1278
QY 1210 ESEKVVQALDKAREGRTCVIAHRLSTIONADLIIVFQNGKVKHEGHTHOOLIAQGIYF 1269
Db 1279 ESEKIVQALDEARKGRTCVIAHRLSTIONSIIAYMSRGIITEQSGHNLMAAKGAYY 1338
QY 1270 SMISVQA 1276
Db 1339 KLVTTGA 1345

RESULT 10
O88331 PRELIMINARY; PRT; 1321 AA.
AC O88331;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE P-GLYCOPROTEIN SISTER.
GN SPGP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxid=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA "Childs S.J., Yeh R.L., Hui D., Ling V.;
RT "taxol resistance mediated by the liver-specific sister gene of p-
RT glycoprotein.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF010597; AAC24753.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; transport.
SQ SEQUENCE 1321 AA; 146286 MW; 27F67EC366008D0F CRC64;
```

Query Match 49.5%; Score 3201; DB 11; Length 1321;
Best Local Similarity 50.2%; Pred. No. 4.2e-177;
Matches 646; Conservative 235; Mismatches 370; Indels 36; Gaps 11;

```
QY 20 KSKKEKKEKKP--TVSTFAMERYSNWLDRLYMLVGTMAAIIHGAALPLMLLVGNMTDS 77
Db 30 KKSRLDKMEGDRLRVGFELFRFSSKDIWLMLMGVCALLHGMQAQGLIIFGIMTDI 89
QY 78 FANAGISR-----NKTFPVIINES-----IYNTQHFHNLHEEMTYYAYYSGI 122
Db 90 FIKYDIERQELETPGKACVNTI-VWINSFHQNMNTGTCGLVDIESEMKFSGIYAGV 148
QY 123 GAGVLAAYIQVSWFCLAGROQLKIRKQFHAIMRQEIQWGFVDVHVGELNLTLDVSK 182
Db 149 GMTVLILGYQIRLWLVITGARIRMRKIYFRIRMRMEIGWFDCTSVGELNSRFADIEK 208
QY 183 INEGIDKVGCMFQSTATFTFGIVGTGPKWKTILVLAISPVLGUSAAIAWILKSFTD 242
Db 209 INDIAIDQLAHFORMSTAMCGLLGFYRGWKLTLVLAISPLIGIGAAYIGLSIAKFT 268
QY 243 KELLAYAKAGAAVEVLAARTVIAFGGKKELERYKNLEAKRTGIGIKKAITANISIGA 302
Db 269 LELKAYAKAGSIADSVLSRTVAAPGGENKEVERTEKNLVFAQRWGIWKMVMVGFTGY 328
QY 303 AFLIIYASALAFWYGTSLVL-SSEYTIQVLTFFSVLIGAFSIGQASPSIEAFANARG 361
Db 329 MWCLIFFCALAFWYGSTLVLDDEEYTPGLVQIFLCVILAAMNIGHASCCLEIFSTGCS 388
QY 362 AAYEIFKIDNKPSIDSYSKSHKPNIKGNLEFKNHFSPSRKEVKILKGLNLKVGSG 421
Db 389 AATNIFQTDROQVIDCMGSDGYKLDRIKGEIEFHNVTFFHPSRPDVKILDNLSMWIKP 448
QY 422 QTVALVNGSGCKSTVQVMORLYDPTDCMVICDGDITINVRHLREITVGVVSOEPLV 481
Db 449 ETTALVNGSGAGKSTALQILQIFDYDCEGWITLDGHDIRSLNIRWLDQIGIVEPEVL 508
QY 482 ATTIAENIRYRENTVMDIEKAVKANAYDFIMLPKNKFDTLVGERGARLSGGQKRIA 541
Db 509 STTIAENIRFREDATMEDIQAAKDANAYNFIMALPQQFDTLVGEQSGQSGQKQV 568
QY 542 IARALVRNPKILLIDDEATSLDTSEAVVOVALDKARKGRTTIVIAHRLSTVRNADVTAG 601
Db 569 IARALVRNPKILLIDDEATSLDSEARVOEALNKIQHGTIISVAHRLSTVRAADVIIG 628
QY 602 FDDGVIVEKGNHDELMKEGIYFKLVMTQTRNGTELENA-TGESKSDALEM----- 655
Db 629 FEHGVAVERTHEELLERKGVYFMLVTLQSQGDNAHETSIMGKDATEGGTLERTFSRG 688
QY 656 PKDSGSLIKRSTR--SIHAP--QGDRKLGTKE-----LNENVPVSFWRIKLIN 705
Db 689 YRDSLRASTRQSKSQLSLTLTHDPLAVADHKSSYKSKDNDVLVEEVEPAPVRILAYN 748
QY 706 STEWYFVVGIFCAINGLQPAFSIIFSRIGITREDDETPETKQNSNMFSLVLGI 765
Db 749 IPEWHYILVGLSAAINGAVTPIYSLFLSQLGTSL-LDKEQQRSEIHSCLFVLGCG 807
QY 766 ISFITFFLQGFQFGKAGEILTKRLRYMVRSMRDVSWFDDPKNTGALTTRLANDAAQ 825
Db 808 VSIFTQFLOGYTFKSGELLTKRLKFGFKAMLGQDQIGWFDLDRNPNVLTIRLATDSQ 867
QY 826 VKGATSLAVITONIANLGTGIIISLYGWQLTLLLAIVPIIAIAGVWEMKMLSGQAL 885
Db 868 VQGATGSGVGMVNSFTNIIAALLIAPFFSWKLSLIITIFFPFLALSGAVQTKMLTGAS 927
QY 886 KDKKELEGAGKIATEAIENFRVTVSLTREQKFEYWAQSLQVPRNSLRKAHIFGVFSFI 945
Db 928 QDQALERAGQITSEALNSIIRTAGIVGEGRIKAFVELOQTSYKTVAKANIYGLCFAF 987
QY 946 TQAMMYFSGCFRGAYLVANEFMNFODVLLVFSIAIVFGAMAYGVSSFPADYAKAKVS 1005
Db 988 SQGTAFLANSAARYGVYLIAYEGLGFSHVFRVSVSVLSATAVGRFTSYTPSTAKAKIS 1047
QY 1006 AAHVIMITEKSPIDSYPHGLKNTLBNVTFNVEVNFNPTRPDIPVLOGLSLEVKGKQ 1065
Db 1048 AARFQLDLDRKPPINVYSEAGEKWDNFQKIDFIDCKFTYPSRPDIQVNLGSLSVNPGQ 1107
```

QY 1066 TIALVSSCGGCTVTVQVLLERYDPLAGSVLIDGKEIKHLNVQWLAHRLHGIIVSQEPILFD 1125
 DB 1108 TIAFVSSCGGCTSTQLLERYDPLAGSVLIDGKEIKHLNVQWLAHRLHGIIVSQEPILFD 1167
 QY 1126 CSTAENIAYDNRVSVSHHEIMQAEANIHHFIETLPEKYNTRYGVKDTQLSGGQKORI 1185
 DB 1168 CSTMDNKKYDGNKEISVERATAAQAQQLHDFVMSLPEKYNTRYGVKDTQLSGGQKORI 1227
 QY 1186 ATARALVROPHILLDEATSDALDTESEKVVQBALDKARGRTCTIVIAHRLSTIQNADLIV 1245
 DB 1228 ATARALVROPHILLDEATSDALDTESEKVVQBALDKARGRTCTIVIAHRLSTIQNADLIV 1287
 QY 1246 VFONGKVKHGTQQLAOKGIYFSMI 1272
 DB 1288 VVSQGVVIEKGTHEKLMKAOAGYKILV 1314

RESULT 11

045721 PRELIMINARY; PRT; 1294 AA.

AC 045721; 062101;
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE C47A10.1 PROTEIN.
 GN C47A10.1.

OC Caenorhabditis elegans.
 OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Basham V.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
 Snelton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Basham V.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; Z93782; CAB07855.1; JOINED.
 DR EMBL; Z81484; CAB07855.1; JOINED.
 DR EMBL; Z81484; CAB03973.1; JOINED.
 DR EMBL; Z93782; CAB03973.1; JOINED.
 DR InterPro; IPR003593; AAA.
 DR InterPro; IPR001140; ABC_transporter_tmem.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR001687; ATP_GTP_A.
 DR InterPro; IPR000130; zn_MTPeptidse.
 DR Pfam; PF00664; ABC_membrane; 2.
 DR Pfam; PF00005; ABC_tran; 2.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
 DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 1294 AA; 142490 MW; DD8ED2602FA484DF CRC64;

Query Match 44.3%; Score 2867; DB 5; Length 1294;

Best Local Similarity 45.7%; Pred. No. 9.8e-158;
 Matches 581; Conservative 239; Mismatches 422; Indels 30; Gaps 9;

QY 19 GKSKKKEKKKPVSTFAMPFRYSNWLDRLVLMVGTWAAIIHGAALPLMLVFGNMTDSF 78
 DB 15 GSSEKKEEAPPPKISIFQILYRTYSTVDRMLAVGIIVSCATGVGLPLMSIINGVSNF 74
 QY 79 ANAGISRNRKTPVITNESITNTTO-----HFINHLEEMTYTAYYSGIGAGVLAAYIQ 133
 DB 75 VTIG-----TIFDPNSTASEKAAAEFSHEVIQNCILKYVY---LGGGIFAAGFLQ 123
 QY 134 VSFWCLAAAGROILKIRKOPFHAIMROEIGFDVHDVDELNRLTDDYDKNEGIDKVG 193
 DB 124 ASCFVAVICEKLSNRFRQFFHVMRQETAWYDKNTSETLSNKLDFNLERVREGTGVKL 183
 QY 194 FQSTATFTFTGTVFGTTPGWLITVILAIAPVLGLSAIWAIIKILSSFTDKELLAYAKA 253
 DB 184 AFQMAAQFIGGFAVAFYDMLTILMSLSFPFMMICGLFLAKLLATAATKAKYAVAG 243
 QY 254 VAEVLAARTVIAFGGQKKELEKRYNKLNEAKRIGIKKAITANISICAPLLIYASYAL 313
 DB 244 IAEVLTISRTVIAFNGOEYCKRYEDALEHGKTKIKKSLFGLAGLASFVVIYASYCL 303
 QY 314 AFWGTSLVLSSEYITGOVLTVFVSVLIGAFSIGOASPSIEAFANARGAAVEIKIDNK 373
 DB 304 AFWGTNFVYSGRESGTVLTVFVSVMMGSMALGAGQGFATIGTALGAAASLIEVDRI 363
 QY 374 PIDSYSGSHKPNKIKGNLEKFNHFSYPSRKEVKILGLNLKLVQSGQVALVNGSGC 433
 DB 364 PEIDAYSTEGQTPSKISGRISVNVKVEFYPTADVILKGVSLDAQPGQOTVALVNGSGC 423
 QY 434 KSTVQLMORLYDPTDGMVCDIGDIRTINVRHLREITGVVSOBPLVFAETIAENIRYGR 493
 DB 424 KSTIQLLQRFYNPDAGQILIDIPEDFNIKYLRQLGVVSOBPLVFAETIAENIRYGR 483
 QY 494 ENVTWDETEKAVKEANAYDFIMKLPKFDTLVGERGLSGGQKORIAIARALVRNPKIL 553
 DB 484 SDVSEDIARALKEANAADFIKTFEGGLTVLGVDRGVMSGQKORIAIARALVRNPKIL 543
 QY 554 LDEATSDALDSEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGDFDGVIVKGNH 613
 DB 544 LDEATSDALDSEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGDFDGVIVKGNH 603
 QY 614 DELMEKGIYKLYVQTRGNEILENATGESKSDALEMSPKDSGLLKRRT--- 670
 DB 604 ETLLIEQGLYHELVAQVFA---DVEDPKKAEARRMSRQTSQRKSGVNFQTSQYDE 660
 QY 671 --RSTHAPOGQDKLG--TKEDLNENVPVSFWRLKLNSTENVPYVGVIFCAIINGLQ 726
 DB 661 KFGAPPAPAEAEKEIKRLKKELEEGAVKANFLKILYARPWEYIFTAIALIQAQVAM 720
 QY 727 PAFSIIFSRIGITRDEDPETKRONSWFVLVLVGLIISITFTFLQGTFGKAGEILT 786
 DB 721 PAFSLFFSQIINVS-NPDRDQMKDGHFWALMFLVLAAVQGTSMFLQCSLFGVAERLT 779
 QY 787 KRLRYMVSRLRODVSFDDPKNTTGALTTLANDAAQVKAIGSLAVITQNIANGT 846
 DB 780 MRIRSKYIRNLVKQDIATYFDMKHSPTITRLATDAPNKAIDYRLGSIENAIASVG 839
 QY 847 GIITSLIYQWLTLLLLAIPIATAGVVMKMSQALKDKKELEGACKIATEIENPR 906
 DB 840 GLGIAFYVQMAFLVMAIFPFMAVQALMKYVHGSATSDAKEMENAGKTAMEAIENR 899
 QY 907 TVVSLTREOKFEYNYAQSLQVYRNSLRKAHIFGVSFSTIQAMVFSYAGCFRFGAYLV- 965
 DB 900 TVQALTLOTKLINIFCSHLDAPHGGNISKAILGLTYGFANSIQFTFYAAAFRFGFLIF 959
 QY 966 -ANEFNFOODLVLSAIVFGAMAVGVQVSFADYAKAVSAAHVIMIIEKSLDISYSP 1024
 DB 960 DKNVLMPEPNLVRLVFAISFCTIGFAASYPEYIKATFAAGLIENMLEEPRIDGMTS 1019
 QY 1025 HGLKPNTLEGNTVFNEVFNYPTRDPILVQLGSLVLEKVGQTALVGVSSGCGKSTVQLL 1084


```

Qy 63 ALPLMLVFGNWTDFANAGISRNKTRPVIIINESITNTQHFINHLEEMT-----YAY 118
Db 64 GFSVLGIVLGMVTVFLRA---QNSEF---VLGTVSRDPEGLPALTKKEEFTLVRRYCLY 117
Qy 119 YSGIGAGVLVAAYTOVFWCLAAAGQILKIRKQFFHAIMRQEGWDFVDVHGEINRLTLD 178
Db 118 YLGIGFAMFATSYIQVCMETFAERITHKRIYKAILRQIQSWFDIQQTGNLTARLTD 177
Qy 179 DVSKINEGDKVGMFQSTATFTPTGTFVGTGPKWKLTLVILAISPVGLSAAIAWAKILS 238
Db 178 DLEVREGDKUSLFTOMVSFAFVGCVFAYSWSMTLVMVYVAFIVISANMMSKIVA 237
Qy 239 SFTDKELLAYAKAGAAVEELAAIRTVIAFGGOKKELERYNKLEAKRIGIKKAITANI 298
Db 238 TRTQVEQETVAVAGAAIEEFTSIRTVHSICGHKRELITFEAALKEGRQGLVKYFMGV 297
Qy 299 SIGNAFLLIYASALAFWYGTSLVSS- EYTGQVLVFFSVLIGAFSGQASPSIEAFA 357
Db 298 GVGFGQCTVSVLAFWYGSVLIINDPALDRGRIETVFAVMGSAALGTCLPHLNTIS 357
Qy 358 NARGAAVEIEFKIIDNKPSIDYSKSGHKPDNIKGNLEFKNVHESYPSRKEVKILGLNLK 417
Db 358 IARGAVSVLSVNSRPKIDPYSLDGVLNNMGRSIRFNKNVHESYPSRRTQLKGVSLQ 417
Qy 418 VQSGQVALVNGSGCKSTTVLMQRLYDPTDGMVCDIGQDRTINVRHLREITGVVSOE 477
Db 418 VSAGQIALVNGSGCKSTVNWLLRFYDPTGKVTIDIDVCDLNVQKIREQIGVSOE 477
Qy 478 PVLFAITIAENIRYGRNTMDIEKAVKEANAYDFIMKLPKFDPLVGERGARGSGOK 537
Db 478 PVLFDGLFENIKMGYEQATMEVQACRVANAADTKRLPEGYGRVGRGVOLSGGOK 537
Qy 538 ORIAIARALVRNPKILLDEATSALDSEAVVOVALDKARKGRTTIVIAHRLSTRVNAD 597
Db 538 ORIAIARAIKNPRILLDEATSALDSEAVVOVALDKARKGRTTIVIAHRLSTRVNAD 597
Qy 598 VIAGFDGIVVEKGNHDEKKEGIVFKLVMTQTRNEIELENATGESKSEDALEMSPK 657
Db 598 QIFVFKNGTIVEQTHAELMKNRGVFEFTMQAQLRKEEVEVLDSADESVVSPDIALP 657
Qy 658 DSGSSLIKRRSTRSIAH-PQGDRLGKTKEDLNENVPVSVFWRILKLNSTWEPFVVG 716
Db 658 HUSSLSRESRISALSVPVSRMGI-EMEDLRAKPTPMS--KIFYFNDRKWGYFILGL 714
Qy 717 FCALINGLOPASFISRIIGFTDEDEPETKRONNMF-SVLFLVLGIISFTIFFLQ 775
Db 715 IACILITGTVTFAVLYAQIQVYS--EPVDMKGHVLFWCAFIIVGLVHAFAPFSA 771
Qy 776 FTFGKAGEILTKRLYVPSRMLRQDVSWDFDPKNTTGALTTRLANDAAQVKAIGSRLA 835
Db 772 ICLGRGEALTKKLRFKFAFNKLLRONVGYDDIRHGTGKLCRTFATDAPNVR-YVFTRLP 830
Qy 836 VITQNIANIGTGIISLIVGWLTLLLAIVPIIAIAGVVEKMLSGOALKDKKELEGAG 895
Db 831 GVLSSVVTIIGALVIGFEGWQALILMVVPLIIGSGTFEMRMQFGKKMRDTELEEGAG 890
Qy 896 KIATEAIENFRVSLTRQKEFEYMAQSLOVPYRNSLRKAHIFGVFSFISITQAMMYFSA 955
Db 891 KVASQAVENIRTVHALNROEPHEFYCEVLEKPEYRENLCQAHTYGVGFASFOSLFFMYA 950
Qy 956 GCFRGAIVLAFNEFNFDVLLVFSALVFGAMVQCVSSFPADYAKAKVSAHVIMITEK 1015
Db 951 VAFWIGALFVDMNHSQPIDVYRVFAFMFGCGQMGVNISSFPDVVVKARLAASLLFYLIEH 1010
Qy 1016 SPLDISYSPHGLKPNLTBEGNTVFNWVFNYPTRPDIPVLOGLSLVKKGTALVGVSGGC 1075
Db 1011 PSEIDNLSDGV-TKKISHSISFRVNYNYPTRQIRVLRGLNLEINPCTTVALVGVSGC 1069
Qy 1076 GKSTVQLLERYDPLAGSVLIDGKEIKHLNVQMLRHLGVQBPILFDCSIAENIAYG 1135
Db 1070 GKSTVALLERFYQNKGVITVDGENIRNMNIRNLREQVCIVSQEPTLFDCTIMENICYG 1129
Qy 1136 DNSRVVSHEEIMQAAKEANIHHFTETLPEKYNTRVGDKGTQLSGQKQRIARALVRQP 1195

```

```

Db 1130 LDDPKFSYEQVAAAKAMANIHNFLVGLPEGYDTRVGEKGTQLSGGQKQRIARALIRDP 1189
Qy 1196 HILLDEATSALDSEKVVQVQALDKARGRTICVIAHRLSTQNALDLVVFQNGKVKREH 1255
Db 1190 PILLDEATSALDSEKIVQDALEVAROGRCTLVIAHRLSTQSDVIVMIOEGKATDR 1249
Qy 1256 GTHQQLAOKQGIYFSMISVQ 1275
Db 1250 GTHELLMKNDLYKRLCETQ 1269

RESULT 15
Q21349
ID Q21349 PRELIMINARY; PRT; 1321 AA.
AC Q21349;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE K08E7.9 PROTEIN.
GN K08E7.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
EMBL; 277666; CAB01232.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1321 AA; 145062 MW; 68094121B30EB746 CRC64;

Query Match 42.48; Score 2740; DB 5; Length 1321;
Best Local Similarity 45.08; Pred. No. 2.4e-150;
Matches 581; Conservative 225; Mismatches 440; Indels 44; Gaps 11;

Qy 20 KKSKEKKEKKPTVSTFAMFRYSNWLDRVLMVGTMAAIIHGAALPLMLVFGNMTDFA 79
Db 47 KITRDAKEEVNKKVSIQLYRYTTTLEKLLFICTIVAVITGAGLPLMSILQGVSAFI 106
Qy 80 NAGISRNKTRPVIIINESITNTQHF-----NHLEEMTVYVYSGIGAGVLVAAY 131
Db 107 NE-----QIVN-----NNGSTFLPTGQNTKTDPEHDMVMNVVSYAAAMTVGMAAGQ 154
Qy 132 IQVSFWCLAAAGRLKIRKQFFHAIMRQEGWDFVDVHGEINRLTDDVSKINEGIDKV 191
Db 155 ITVTCYLVVAEQMNNRLRREFVKSILRQELSWDTNHSIGTFLATKLFNLREREGTGDKI 214
Qy 192 GMFQSTATFTGTFVGTGPKWKLTLVILAISPVGLSAAIAWAKILSSFTDKELLAYAKA 251
Db 215 GMAFYLSQFTGTGTFVAFTHSWQLTVMLAVTPTQALCGFAIAKSKSTFAIRETLAYAKA 274
Qy 252 GAAVEELAAIRTVIAFGGOKKELERYNKLEAKRIGIKKAITANISGAAFLIYASY 311
Db 275 GKVVEETISSIRTVSVSLNGLRYELERYSTAVEAKKAGVLKGLFLGISFGAMQASNFISF 334

```

QY 312 ALAFWYGTSLVLSSEYTIQOVLTFFSVLIGAFSIGQASPSIEAFANARGAAEYFKIID 371
| | | | : | : | | | | : : | | : | | | | : : : |
Db 335 ALAFYIGVGWHDGSLNFGDMITTFSSVMGSMALAGLQPLAVLGTAGGRASGYEYVLD 394
| | | | : | : | | | | : : | | : | | | | : : : |
QY 372 NKPSIDSYSKGKHDPDNKGNLEKFNHFSYPSRKEVKILKGLNLKVQSGOTVALVGNSSG 431
| | | | : | : | | | | : : | | : | | | | : : : |
Db 395 RKPVIDSSSKAGRKQMKIKGDITVENVHFTYPSRDPVPIIRGMNLRVNAGOTVALVSSG 454
| | | | : | : | | | | : : | | : | | | | : : : |
QY 432 CGKSTTVQLMQRLYDPTDGMWIDGODIRTIIVRHRLREITGVVSOBPVLFATTIAENIRY 491
| | | | : | : | | | | : : | | : | | | | : : : |
Db 455 CGKSTIISLLRYDYVLKGITIDGVDRDINLEFLRKNVAVVQBPALFNCITIEENISL 514
| | | | : | : | | | | : : | | : | | | | : : : |
QY 492 GRENTMDETEKAVKEANAYDTMKLPNKFDTLVGERGARLSGGQKQRIATARALVRNPK 551
| | | | : | : | | | | : : | | : | | | | : : : |
Db 515 GREGITREEMVAVACKMANAEKIKTLPNGYNTLVGDRGTQLSGQKQRIATARALVRNPK 574
| | | | : | : | | | | : : | | : | | | | : : : |
QY 552 ILLDDEATSALDTESEAVVQVALDKARKORTTIVIAHRLSTVRNADVITAGDDGVIVEKG 611
| | | | : | : | | | | : : | | : | | | | : : : |
Db 575 ILLDDEATSALDAESGIVQALDKAAGRTTIIAHRLSTIRNADLIISCKNGQVVEYG 634
| | | | : | : | | | | : : | | : | | | | : : : |
QY 612 NHDELMEKEGIYFKLYTMOTRGNEIE- - - - -LENATGESKESDALEMSPKDSGSSL 663
| | | | : | : | | | | : : | | : | | | | : : : |
Db 635 DHRALMAQOGLYYDLVTAQTTDAVDSAAGKFSRENSVAROTSEHEGLSRQASEMDDIM 694
| | | | : | : | | | | : : | | : | | | | : : : |
QY 664 IK-RRSTRRI-HAQO- - - - -GQDRKLGTKEDLNE-NVPSVSFWRLKLNSTWPFYF 712
| | | | : | : | | | | : : | | : | | | | : : : |
Db 695 NRVRSSTIGSITNGPVIDEKEERIGKDALSRUKQELEENNAQKTNLFELYHARPHALS 754
| | | | : | : | | | | : : | | : | | | | : : : |
QY 713 VGFICAITNGGLQAPAFSIFRSIIIGITRDEDPETKRONSNMFSVLFLVLGIISFIFF 772
| | | | : | : | | | | : : | | : | | | | : : : |
Db 755 FIGMSTATIGGFIYPTYSVFTSFNMVFA--GNPADELQGHFWALMFLVLAAAGQICSF 812
| | | | : | : | | | | : : | | : | | | | : : : |
QY 773 LQGFTECKAGEIILTKRLRVNFSMLRQDVSFDDPKNTTGALTTRLANDAAQVKGAGS 832
| | | | : | : | | | | : : | | : | | | | : : : |
Db 813 LMTFFMGIASESITRDNKLFNLSQHIGFEDSQNASGKISTRLATDVNLRITADF 872
| | | | : | : | | | | : : | | : | | | | : : : |
QY 833 RLAVITONIANLGTIISLIYGWQLTLLLAIVPIIAIAGVVMKMLSGQALKKKKELE 892
| | | | : | : | | | | : : | | : | | | | : : : |
Db 873 RFSVTITLVSMVAGIGLAFYQWQALLIIAILPIVAFQYLRGRFTGKNVKSASEFA 932
| | | | : | : | | | | : : | | : | | | | : : : |
QY 893 GAGKIATEAIENTFVVSITREQFEYMYAQSLQVYRNSLRKAHIFGVYSFISITQAMMYF 952
| | | | : | : | | | | : : | | : | | | | : : : |
Db 933 DSGKIAIEAIENTVTOALAREDTFENFCEKLDIPHKEAIEAFIOGLSYGCASSVLYL 992
| | | | : | : | | | | : : | | : | | | | : : : |
QY 953 SYACGFRFGAYLVANEFMNFQD--VLIVFSAIVFGAMAVGVQSVSFAPDYAKAKVSAHVI 1010
| | | | : | : | | | | : : | | : | | | | : : : |
Db 993 LNTCAYRMGLALIITDPTWQPMRVLNVMYAITISTTLGFATSYPEYAKATFAGGIIF 1052
| | | | : | : | | | | : : | | : | | | | : : : |
QY 1011 MIEKSPIDISYSPHGLKPNLTGNVTNFVFNYPTRPDIPVLOGLSLEVKKGOTLALV 1070
| | | | : | : | | | | : : | | : | | | | : : : |
Db 1053 GMLRKISIDSLSLAGEK-KLYGKVIKFNVRFAYPEPERPEIILKGLSFSEVPGQTALV 1111
| | | | : | : | | | | : : | | : | | | | : : : |
QY 1071 GSSCGKSTVVOLLERYDPLAGSVLIDGKEIKHLNQMVLRAHLGIVSQBPILFDCSIAE 1130
| | | | : | : | | | | : : | | : | | | | : : : |
Db 1112 GPSGCGKSTVVALLERYDYLGGEIFIDGSEIKTLNPEHTRSQIAIVSQEPTLFDCSIAE 1171
| | | | : | : | | | | : : | | : | | | | : : : |
QY 1131 NIAYGDNRVVSHEIMQAAKEANIHFIETLPKYNTRVGDGTQLSGGQKQRIARA 1190
| | | | : | : | | | | : : | | : | | | | : : : |
Db 1172 NIIVGLDPSVTVMAQVEAARLANIHNFIAELPEGFETRVGDRGTQLSGGQKQRIARA 1231
| | | | : | : | | | | : : | | : | | | | : : : |
QY 1191 LVROPHILLDDEATSALDTESEKVVQALDKAREGRTCIIVIAHRLSTIONADLIVVFONG 1250
| | | | : | : | | | | : : | | : | | | | : : : |
Db 1232 LVARNPKILLDDEATSALDTESEKVVQALDKAREGRTCIIVIAHRLNIVMNADCIANVSNG 1291
| | | | : | : | | | | : : | | : | | | | : : : |
QY 1251 KVKEHGTHTQQLAOKGIYFMSISVQAGAKR 1280
| | | | : | : | | | | : : | | : | | | | : : : |
Db 1292 TIEKGHTHTQLMSEKAGYKLTQKQMTTEK 1321
| | | | : | : | | | | : : | | : | | | | : : : |

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 19:26:40 ; Search time 2229.22 Seconds
(without alignments)
7749,846 Million cell updates/sec

Title: US-09-672-725C-4

Perfect score: 6467

Sequence: 1 MDPGKRGKSAENFWKMGK.....LLAQKIYFMSISVQAGAKR 1280

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09672725/runat_04112002_124342_18001/app_query.fasta_1.6435
-DB=EST -Qfmt=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blossum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725 @CGL_1_7365 @runat_04112002_124342_18001 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1420	22.0	943	9	AL520322
2	1073.5	16.6	981	10	BF969667

3	1038	16.1	803	10	BG298756
4	1031	15.9	894	10	BF584668
5	1024	15.8	1019	10	BG248052
6	968.5	15.0	944	10	BF796582
7	958	14.8	851	10	BF969062
8	950	14.7	785	9	AV709991
9	916.5	14.2	795	10	BF313560
10	916	14.2	726	10	BG293345
11	910	14.1	587	10	BG293345
12	908.5	14.0	628	10	BM053000
13	889	13.7	563	10	BG080311
14	889	13.7	563	10	BF964198
15	889	13.7	563	10	BF964218
16	880	13.6	605	9	BM052878
17	876	13.5	2676	11	AK014319
18	875	13.5	563	10	BM053038
19	826	12.8	546	10	BF964137
20	821.5	12.7	1201	10	BF304387
21	806.5	12.5	871	12	AZ682350
22	805	12.4	759	10	BE376459
23	792	12.2	662	10	BG963638
24	790.5	12.2	886	12	BH139685
25	780	12.1	517	10	BF964115
26	777.5	12.0	897	12	AZ541090
27	768	11.9	780	10	BG587938
28	766.5	11.9	932	12	AZ670821
29	764	11.8	560	10	BE749379
30	763.5	11.8	834	12	AZ548312
31	762.5	11.8	913	12	BH155700
32	762	11.8	822	10	B1332761
33	761.5	11.8	947	12	AZ683753
34	758.5	11.7	880	12	AZ687805
35	756	11.7	669	10	BM013981
36	749.5	11.6	823	12	AZ532602
37	748.5	11.6	853	12	AZ679807
38	747.5	11.6	919	12	AZ680701
39	744.5	11.5	511	9	AA239727
40	743.5	11.5	886	12	AZ540627
41	742	11.5	600	10	BM486593
42	741.5	11.5	891	12	AZ682250
43	737	11.4	699	9	BB667773
44	730	11.3	760	10	BM016204
45	729.5	11.3	899	12	BH154857

ALIGNMENTS

AL520322 943 bp mRNA linear EST 13-FEB-2001
AL520322 LTI_NFL004_NBC2 Homo sapiens cdna clone CS0DB006YC15 5
prime, mRNA sequence.
AL520322 GI:12783815
EST.. human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 943)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .943
/organism="Homo sapiens"
/db_xref="taxon.9606"
/clone="CS0DB006YC15"
/clone_lib="LTI_NFL004_NBC2"

```

/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DHI0B"
/Note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 253 a 225 c 253 g 211 t 1 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1.76e-120 Length: 943
Score: 1420.00 Matches: 282
Percent Similarity: 96.41% Conservativeness: 13
Best Local Similarity: 92.16% Mismatches: 10
Query Match: 21.96% Indels: 1
DB: 9 Gaps: 0

US-09-672-725c-4 (1-1280) x AL520322 (1-943)

Qy 976 LeuLeuValPheSerAlaIleValPheGlyAlaMetAlaValGlyGlnValSerSerPhe 995
Db 2 CTGTTATTATTTCAGCTGTGTCTTTGGTGGCATGGCGTGGGCAAGTCAGTTCATTT 61

Qy 996 AlaProAspTyrAlaLysAlaLysValSerAlaAlaHisValMetIleLeuGlyLys 1015
Db 62 GTCCTGACATATGCCAAGCAAAATATCAGCAGCCCATCATCATCATGTAAGAAA 121

Qy 1016 SerProLeuLeuAspSerTyrSerProHisGlyLeuLysProHisThrLeuGlyGln 1035
Db 122 ACCCTTTGATTCAGCTACAGCAGGAGGAGGCTAATGCCGAACATTTGGAAGAAAT 181

Qy 1036 ValThrPheAsnGlnValValPheAsnTyrProThrArgProAspIleProValLeuGln 1055
Db 182 GTCACATTTGGTGAAGTTGTATTCACACTATCCACCCGACCCGACATCCAGTGCCTCAG 241

Qy 1056 GlyLeuSerLeuGlnValLysLysGlyGlnThrLeuAlaLeuValGlySerGlyCys 1075
Db 242 GGACTGAGCTGGAGGTGAAGAGGCGCAGACGCTGCTGCTGGTGGGCGACGAGTGGCTGT 301

Qy 1076 GlyLysSerThrValValGlnLeuLeuGluArgPheTyrAspProLeuAlaGlySerVal 1095
Db 302 GGGAGAGCAGCAGGTGTCAGAGGCGCAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361

Qy 1096 LeuLeuAspGlyLysGlnLysHisLeuAsnValGlnTrpLeuArgAlaHisLeuGly 1115
Db 362 CTGCTTGTGCAAGAAATTAAGCGACTGAATGTTTCAGTGGCTCCGAGCAGACCTGGGC 421

Qy 1116 IleValSerGlnGluProIleLeuPheAspCysSerIleAlaGlnAsnIleAlaTyrGly 1135
Db 422 ATCGTGTCCAGGAGCCATCTCTTTTGTACTGACGACATTCGAGAACATTCGCTATGGA 481

Qy 1136 AspAsnSerArgValValSerHisGlnGluIleMetGlnAlaLysGluAlaAsnIle 1155
Db 482 GACAAACAGCCGGGTGGTGTACAGAGAGAGATCGTAGGCGCAGCAAGAGGCGCACATA 541

Qy 1156 HisHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThr 1175
Db 542 CATGCTTCATCGAGTCACTGCCTTAATAATATAGCACTAAAGTAGGAGACAAGGAAC 601

Qy 1176 GlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnPro 1195
Db 602 CAGCTCTGTGTGGCCAGAAACAGCATTCGCCATGCTGTGCCCTTGTAGACACCT 661

Qy 1196 HisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysValVal 1215
Db 662 CATATTTTGGTGGTGAAGCCCGTCAGCTCTGGGATACAGAAAGTGAAGAGTTGTC 721

```

```

Qy 1216 GlnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLeu 1235
Db 722 CAAGAAGCCCTGGACAAAGCCAGAGAGCGCCACCTGCATTGTGATTGCTCACCAGCTG 781

Qy 1236 SerThrIleGlnAsnAlaAspLeuIleVal-ValPheGlnAsnGlyLysValLysGluHi 1255
Db 782 TCCACCATCCAGAAATGCAGACTTAATAGTGTGTGTTTCAGAAATGGCAGAGCAAGGAGCA 841

Qy 1255 sGlyThrHisGlnLeuLeuAlaGlnLysGlyIleTyrPheSerMetIleSerValG1 1275
Db 842 TGGCAGCATCAGCAGCTGCTGGCAGAGAAAGGCAATCTATTTTCAATGTCAGTGTCCA 901

Qy 1275 nAlaGlyAlaLysArg 1280
Db 902 GGCTGGRACAAAGCGC 917

RESULT 2
BF969667 981 bp mRNA linear EST 22-JAN-2001
LOCUS 602272046F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4360090 5',
DEFINITION mRNA sequence.
ACCESSION BF969667
VERSION BF969667.1 GI:12336882
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 981)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10000 row: 1 column: 11
High quality sequence stop: 670.
FEATURES
source
1. 981
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4360090"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DHI0B (phage-resistant)"
/Note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 267 a 194 c 250 g 270 t
ORIGIN
Alignment Scores:
Pred. No.: 1.58e-88 Length: 981
Score: 1073.50 Matches: 234
Percent Similarity: 86.60% Conservativeness: 18
Best Local Similarity: 80.41% Mismatches: 31
Query Match: 16.60% Indels: 10
DB: 10 Gaps: 5

US-09-672-725c-4 (1-1280) x BF969667 (1-981)

Qy 792 MetValPheArgSerMetLeuArgGlnAspValSerTyrPheAspProLysAsnThr 811
Db 1 ATGGTTTTCGATCCATGCTCAGACAGAGATGTGAGTTGTTGTCACCTAAACACACC 60

```


QY	415	AsnLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGlyLys	434
Db	541	AAATCTGAAGGTGAAGAGCGGACAGACGGTGGCTGGTGGCAACAGTGGCTGGGAAAA	600
QY	435	SerThrValGlnLeuMetGlnArgLeuThrAspProThrAspGlyMetValCysIle	454
Db	601	AGCAACAACCTGCTCCAGCTGATGCAAGGCTCTACGACCCCTA-GATGGCATGGTCAGTATC	659
QY	455	AspGlyGlnAspIleArgThrIle-AsnValArgHisLeuArgGluIleThrGlyValVa	474
Db	660	GACGGACAGACATCAGAACCATCAATGTGAGGTATCTGGAGGGGATCATGGGGTGGT	719
QY	474	lSerGlnGluProValLeuPheAlaThrIleAlaGluAsnIleArgThrGly	492
Db	720	AGTCAAGGAACCTGGCTGTTGTCACCAACCCAGATCGCGAGAAAATTCGTTATGGG	774
RESULT 4			
BF584668		894 bp mRNA linear EST 12-DEC-2000	
LOCUS	602098406F1	NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4218385 5',	
DEFINITION		mRNA sequence.	
ACCESSION	BF584668		
VERSION	BF584668.1	GI:11658386	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 894)		
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: Jeffrey E. Green, M.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LLAM9796 row: 1 column: 02		
	High quality sequence stop: 651.		
FEATURES	Location/Qualifiers		
source	1..894		
	/organism="Mus musculus"		
	/strain="FVB/N"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:4218385"		
	/lab_host="NCI_CGAP_Co24"		
	/notes="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;		
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.		
	Average insert size 1.6 kb. Constructed by Life		
	Technologies. Note: this is a NCI_CGAP Library."		
BASE COUNT	263 a 203 c 242 g 185 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1..16e-84	Length:	894
Score:	1031.00	Matches:	224
Percent Similarity:	92.48%	Conservative:	22
Best Local Similarity:	84.21%	Mismatches:	15
Query Match:	15..94%	Indels:	9
DB:	10	Gaps:	1
US-09-672-725C-4 (1-1280) x BF584668 (1-894)			
QY	334	ThrValPheSerValLeuIleGlyAlaPheSerIleGlyGlnAlaSerProSerIle	353
Db	2	ACTGCTCTCTTTCGCTGTTAAT-GGAGCATTCAGTCTGGCAGGCGATCTCCAAATAT	60
QY	354	GluAlaPheAlaAsnAlaArgGlyAlaAlaTyrGluIlePheIleIleAspAsnLys	373
Db	61	GAAGCCTTCGCAATGACGAGGAGCAGCTTATGAGTCTTCAAAATAATTGATAATAAG	120
QY	374	ProSerIleAspSerTyrSerLysSerGlyHisLysProAspAsnIleLysGlyAsnLeu	393
Db	121	CCCAATATAGACAGCTTCTCAAGAGTGGGCACAAACACCAACATACAAAGAAATCTG	180
QY	394	GluPheLysAsnValHisPheSerTyrProSerArgLysGluValLysIleLeuLysGly	413
Db	181	GAATTTAGAATAATTCATCTCAGTTACCCATCTCGAAAGAAGTTCAGATCTTTGAAGGC	240
QY	414	LeuAsnLeuLysValGlnSerGlyGlnThrValAlaLeuValGlyAsnSerGlyCysGly	433
Db	241	CTCAATCTGAAGGTGAAGAGCGGACAGCGGTGGCTGGTGGCAACATGCTGTGGA	300
QY	434	LysSerThrThrValGlnLeuMetGlnArgLeuTyrAspProThrAspGlyMetValCys	453
Db	301	AAAGCAACAACCTGCTCCAGCTGATGCAAGGCTCTACGACCCCTAGATGCGTGCAGT	360
QY	454	IleAspGlyGlnAspIleArgThrIleAsnValArgHisLeuArgGluIleThrGlyVal	473
Db	361	ATCGACGGACAGGACATCAGAACCATCAATGTGAGGTATCTGAGGGAGATCATTTGGTGTG	420
QY	474	ValSerGlnGluProValLeuPheAlaThrThrIleAlaGluAsnIleArgTyrGlyArg	493
Db	421	GTGAGTCAAGAAACCTGCTGCTGTTGCCACACGATCGCCGAGAACATCGCTATGGCCGA	480
QY	494	GluAsnValThrMetAspGluIleGlyLysAlaValLysGluAlaAsnAlaTyrAspPhe	513
Db	481	GAAGATGTCACCATCGATGAGATTGAGAAAGCTGTCAAGGAAGCAATGCTATGACTTC	540
QY	514	IleMetLysLeuProAsnLysPheAspThrLeuValGlyGluArgGlyAlaArgLeuSer	533
Db	541	ATCATGAAACTGCCCCACCAATTTGACACCTTGGTGGTGGAGAGGGCCGACCTGAGT	600
QY	534	GlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgAsnProLysIleLeu	553
Db	601	GGGGGACAGAAACAGAGATCGC-ATTGCCGGGCCCTGGTCCGCAATCCCAAGTGCTT	659
QY	554	LeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAla	573
Db	660	TC-GTGGACGAGCCACCTCAGC-CTGGATACAGAAAGTGAAGCTGGGGTTCAGCCGCCA	717
QY	574	-LeuAspLysAlaArgLysGlyArgThrThrIle-----ValIleAlaHisAr	589
Db	718	CCTAGATAGGTGTCAAAGAGGCGNGGACCATTTGGGAAGCTCATCGCTTTACCGTCGAA	777
QY	589	gLeuSerThrValArg	594
Db	778	GCTGACGGCCATGCGG	793
RESULT 5			
LOCUS	60235987F1	NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4488404 5',	
DEFINITION		mRNA sequence.	
ACCESSION	BG248052		
VERSION	BG248052		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 1019)		
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/.		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: Gilbert Smith, Ph.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10334 row: n column: 21
 High quality sequence stop: 650.
 Location/Qualifiers
 1..1019

FEATURES

source
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="INAGE:4488404"
 /clone_lib="NCI_CGAP_Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="10 months, virgin"
 /lab_host="DH108"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
 BASE COUNT 276 a 262 c 283 g 196 t 2 others
 ORIGIN

Alignment Scores:

Pred. No.: 6.05e-84 Length: 1019
 Score: 1024.00 Matches: 216
 Percent Similarity: 90.48% Conservative: 12
 Best Local Similarity: 85.71% Mismatches: 19
 Query Match: 15.83% Indels: 5
 DB: 10 Gaps: 2

US-09-672-725C-4 (1-1280) x BG248052 (1-1019)

Qy 1032 LeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArgProAspIle 1051
 Db 1 TTGGAAGCAATGTGCAATTAGTGGAGTCTGTTCAACTATCCACCCAGCCACATC 60
 Qy 1052 ProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGly 1071
 Db 61 CCAGTGTCTCAGGGCTGAGCCTTGAGGTGAAGAGGCGCAGCGTGGCCCTGGTGGC 120
 Qy 1072 SerSerGlyCysGlyLysSerThrValValGlnLeuGluArgPheThrAspProLeu 1091
 Db 121 AGCAGTGGCTGGGGAAGACACAGTGTCCAGTCTCGAGCGCTTACACCCCATG 180
 Qy 1092 AlaGlySerValLeuLeuAspGlyLysGluIleLysHisLeuAsnValGlnThrLeuArg 1111
 Db 181 GCTGGATCAGTGTCTAGATGGCAAGAAATAAACAACCTGAATGTCAGTGGCTCGA 240
 Qy 1112 AlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsn 1131
 Db 241 GCACAGCTGGGCATTGTGCCAAGAGAGCCCATCTCTTTGACTGCGAGCATCGCAGA 300
 Qy 1132 IleAlaTyrGlyAspAsnSerArgValValSerHisGluGluIleMetGlnAlaAlaLys 1151
 Db 301 ATTGCCTACGAGACACAGCGGTGCTGTTATGAGGAGATTGTGAGGCGACCCAAG 360
 Qy 1152 GluAlaAsnIleHisHisPheLeuThrLeuProGluLysTyrAsnThrArgValGly 1171
 Db 361 GAGGCCAACATCCACGAGTTCATCGACTCGCTACCTGATAAATAACAACACAGAGTAGA 420
 Qy 1172 AspLysGlyThrGlnLeuSerGlyGlnGlnLysGlnArgIleAlaIleAlaArgAlaLeu 1191
 Db 421 GACAAAGGCACATCAGCTGCGGGTGGGAGAGAGCGCATCGCCATCGCAGCGCCCTC 480
 Qy 1192 ValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer 1211
 Db 481 GTCAGACAGCCTCACATTTTACTTCTGGAGGAGCAACATCAGCTCTGGATACAGAAGT 540
 Qy 1212 GluLysValValGlnGlnAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIle 1231
 Db 541 GAAAGGTTGTCCAGGAGCGTGGACAAAGCCAGGAAGCCGACCTGCATGTGTATC 600
 Qy 1232 AlaHisArgLeuSerThrIleGlnAsnAlaAspLeuIleValVal-PheGlnAsnGlyL 1251

Db 601 GCTCACCCTGTCCACCATCCCAAGCCGACTGTGCTGGTGTGATTCAGAACCGGA 660
 Qy 1251 ysValLys---GluHisGlyThrHisGlnGlnLeuLeuAlaGlnLysGlyIleTyrPhe 1270
 Db 661 AGGTCAAGGAGCAGCGGCACCCCAACAGCAAGTGGTGGCGCAGAGGGCATCTACTTCT 720
 Qy 1270 erMetIleSer-ValGlnAlaGlyAlaLys 1279
 Db 721 CCAGG---TCAGGTGCATGCTGCGCCCAAG 747
 RESULT 6
 LOCUS BF796582
 DEFINITION 602258463p1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4341710 5',
 mRNA sequence.
 ACCESSION BF796582
 VERSION BF796582.1 GI:12101636
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 944)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Louis Staudt, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM9955 row: n column: 15
 High quality sequence start: 9
 High quality sequence stop: 669.
 Location/Qualifiers
 1..944
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4341710"
 /clone_lib="NIH_MGC_85"
 /tissue_type="lymphoma, cell line"
 /lab_host="DH108 (phage-resistant)"
 /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT

265 a 241 c 237 g 201 t

ORIGIN

Alignment Scores:

Pred. No.: 7.15e-79 Length: 944
 Score: 968.50 Matches: 210
 Percent Similarity: 83.64% Conservative: 20
 Best Local Similarity: 76.36% Mismatches: 39
 Query Match: 14.98% Indels: 8
 DB: 10 Gaps: 1

US-09-672-725C-4 (1-1280) x BF796582 (1-944)

Qy 987 MetaValGlyGlnValSerPheAlaPro-Asp-TyrAlaLysAlaLysValSera 1006
 Db 1 GTGGCTCTAGGACATGCCAGTTCATTTGCTCCAGGACGTATGCTAAAGCTAAGCTGCTG 60
 Qy 1006 laAlaHisValIleMetIleIleGluLysSerProLeuIleAspSerTyrSerProHisg 1026
 Db 61 CAGCCCATTTATTTCATGCTGCTTTGAAGACACACCTCTGATTGACAGCTACAGTGAAGAG 120

QY 1026 lyLeuLysProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrP 1046
 Db 121 GGCTGAAGCTGATAAATTTGAAGGAATATACATTTAATGAAGTCGGTTCACATATC 180
 QY 1046 roThrArgProAspIleProValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnT 1066
 Db 181 CCACCCGAGCAACGGTCCAGCTCTTCAGGGCTGAGCTGGAGGTGAAGAAGGCCAGA 240
 QY 1066 hrLeuAlaLeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluA 1086
 Db 241 CACTAGCCCTGGTGGCAGCAGTGGCTTCAGGGCTGAGCTGGAGGTGAAGAAGGCCAGA 300
 QY 1086 rgPheTyrAspProLeuAlaGlySerValLeuLeuAspGlyLysGlnLeuLysHisLeuA 1106
 Db 301 GGTCTACGACCCCTTGGCGGGGACAGTCTCTCATGTGTCAAGAGGCAAGAACTCA 360
 QY 1106 snValGlnTTrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspC 1126
 Db 361 ATGTCCAGTGGCTCAGAGCTCAACTCGGAATCGTGTCTCAGGAGCCTATCTATTGACT 420
 QY 1126 ysSerIleAlaGluAsnIleAlaTyrGlyAspAsnSerArgValValSerHisGluGlnI 1146
 Db 421 GCAGCATTTGGCGGAGAAATTTGCTATGGAGACAAACGCGGTGTATCATCAGAGTAA 480
 QY 1146 leMetGlnAlaAlaLysGluAlaAsnIleHisPheIleGluThrLeuProGluLysT 1166
 Db 481 TCGTGAGTGCAGCCAAAGCTGCCAATACATCTCTTCATCGAGAGCTTACCCCAAAAT 540
 QY 1166 yrAsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleA 1186
 Db 541 ATGAACAAGAGTGGAGATAAGGGGACTCAGCTCTCAGGAGGTCAAAACAGAGGATTG 600
 QY 1186 laIleAlaArgAlaLeuValArgGln-ProHisIleLeuLeuLeuAspGluAlaThrSer 1205
 Db 601 CTATGCGCGGACCTCATAGACAAACCCCAATCTCTCTGTGGTGAAGTACATCA 660
 QY 1206 Ala-LeuAspThr-----GluserGluLysValValGlnGluAlaLeuAspLysAla 1223
 Db 661 GCCTCTGATACCTGCACAGTGAACACGGTGTGCCAAGAAGCCCTGTCCACAGCAGA 720
 QY 1223 gLcGlyArgThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAlaAspLe 1243
 Db 721 AGAGCGCGCACCTCGCTTGTGATGTCTACCGCTGTC--AACATCAGGAATGCAGACT 778
 QY 1243 urLeuValPheGlnAsnGlyLysValLysGluHis 1255
 Db 779 -ATAGTGTGTTCACACGGGAGAGTCAACACACAT 814
 RESULT 7
 BF969062
 LOCUS 1 (bases 1 to 851)
 DEFINITION 60226993f1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4358135 5',
 mRNA sequence.
 ACCESSION BF969062
 VERSION BF969062.1 GI:12336277
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 851)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMLL at:
 http://image.llnl.gov

Plate: L1AM9995 row: j column: 24
 High quality sequence stop: 612.
 Location/Qualifiers
 source 1..851
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4358135"
 /clone_lib="NIH_MGC_84"
 /tissue_type="adrenal cortex carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site: 1:
 NotI; Site: 2: SalI; Cloned unidirectionally; oligo-dr
 primed. Average insert size 1.229 Kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 243 a 172 c 214 g 222 t
 ORIGIN
 Alignment Scores: 5.81e-78 Length: 851
 Pred. No.: 217
 Score: 958.00 Matches: 217
 Percent Similarity: 79.25% Conservative: 16
 Best Local Similarity: 73.81% Mismatches: 33
 Query Match: 14.81% Indels: 29
 DB: 10 Gaps: 5
 US-09-672-725C-4 (1-1280) x BF969062 (1-851)
 QY 825 GlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeu 844
 Db 3 CAAGTTAAGGGGCTATAGTTCAGGCTTCTGTAATATACCCAGATATAGCAATCTT 62
 QY 845 GlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAla 864
 Db 63 GGCACAGGAATAATTATATCTTCATCTATGTTGGCACTAACACTGTTACTCTTAGCA 122
 QY 865 IleValProIleIleAlaIleAlaGlyValValGluMetLysMetLeuSerGlyGlnAla 884
 Db 123 ATTGTACCATCATTCGAATAGCAGGAGTGTGTAATGAAATGTAATGTAATGTAATGTA 182
 QY 885 LeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsn 904
 Db 183 CTGAAGATAAAGAAAGAACTAGAGGTGCTGGGAAGATCGCTACTGAGCAATAGAAAC 242
 QY 905 PheArgThrValValSerLeuThrArgGluGlnLysPheGluTyrMetTyrAlaGlnSer 924
 Db 243 TTCGGAACCGCTGTTCTTTGACTCAGGACAGAAAGTTGAACATATATGATGCTCAGAGT 302
 QY 925 LeuGlnValProTyrArgAsnSerLeuArgLysAlaHisIlePheGlyValSerPheSer 944
 Db 303 TTCAGGTACCATACAGAAACCTTTGAGAAAGCACACATCTTTGGAATTACATTTTCC 362
 QY 945 IleThrGlnAlaMetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeu 964
 Db 363 TTCACCCAGGCAATGATGATTTTCTCTATGCTGGATGTTCCGGTTTCGAGCCTACTTG 422
 QY 965 ValAlaAsnGluPheMetAsnPheGlnAspValLeuLeuValPheSerAlaIleValPhe 984
 Db 423 GTGGCACAATACTCATGAGCTTTGAGGATGTTCTGTAGTATTTTTCAGCTGTGCTTT 482
 QY 985 GlyAlaMet-AlaValGlyGlnValSerPheAlaProAspTyrAlaLysAlaLysVa 1004
 Db 483 GTTGCCATGGCGCTGGGCAAGTCAGTTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 542
 QY 1004 lSerAlaAlaHisValIleMetIleIleGluLysSerProLeuIleAspSerTyrSerPr 1024
 Db 543 ATCAGCAGCCCATCATCATGATCATTTGAAAAAACCCTTTTAT-GACAGCTACAGCAC 601
 QY 1024 oHisGlyLeuLysProAsnThr---LeuGluGlyAsnValThrPhe-----AsnGluVa 1041
 Db 602 GGAAGGCGCTAATTCGCCGGAACACATTTGGGAAGGAATGTACCATGTGGTGAACCTGT 661
 QY 1041 lValPheAsnTyrProThrArg-ProAspIleProValLeuGlnGlyLeuSerLeuGluV 1061

```

Db 662 TTTTCAAAATTTCAACCCCTAGCGACATCCAGTGCCTCCAGGACTGAAGCTGGGT 721
QY 1061 aLysLysGlyGlnThrLeuAlaLeu-----ValGlySerSer----- 1073
Db 722 GAAGAAAGGGGCAAAAGCGGGTCTCGGTGGGCACCAATGGTGGTGGGAAAGACACCA 781
QY 1074 -----GlyCysGlyLysSerThrValGlnLeuLeuGluAlaGPhyTyra 1089
Db 782 GGGGCCAGTCCCGGAACCGGTCTAAAC----- 813
QY 1089 spProLeuAlaGlySerValLeuLeuAspGlyLysGlu 1101
Db 814 -----CTGGCGGAATGCTGCTCTGTGGGCAAAAG 846

RESULT 8
AV709991 LOCUS 785 bp mRNA linear EST 09-OCT-2000
DEFINITION AV709991 Cu Homo sapiens cDNA clone CUAJA09 5', mRNA sequence.
ACCESSION AV709991
VERSION AV709991.1 GI:10728272
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 785)
AUTHORS Peng, Y., Song, H., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N.,
Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu
W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J.,
Chen, Z. and Han, Z.
TITLE Homo sapiens cDNA Cu clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@hgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CUAJA09"
/clone_lib="Cu"
/tissue_type="adrenal cortico adenoma for Cushing's
syndrome"
/dev_stage="Adult"
/note="Vector: pBluescript sk(-)"
BASE COUNT 210 a 180 c 211 g 183 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 2,86e-77 Length: 785
Score: 950.00 Matches: 208
Percent Similarity: 86.87% Conservative: 17
Best Local Similarity: 80.31% Mismatches: 33
Query Match: 14.69% Indels: 4
DB: 9 Gaps: 0

US-09-672-725c-4 (1-1280) x AV709991 (1-785)

QY 954 TyraLacGlyCysPheArgPheGlyAlaTyriLeuValAlaAsnGluPheMetAsnPheGln 973
Db 13 TACTAAGGGCTGCAAGCTTTGGCAATCCAGGATATGTTACTTCCCTTGATAAGAGA 72
QY 974 AspValLeuValPheSerAlaLeuValPheGlyAlaMetAlaValGlyGlnValSer 993
Db 73 GGAATTGGAAGAGATATTTTCAGCTGTTGCTTTGGTGCCATGGCGGGGCAAGTCAGT 132
QY 994 SerPheAlaProAspTyraLalysAlalysValSerAlaAlaHisValIleMetIle 1013

```

```

Db 133 TCATTGTCTCTGACTATGCAAAATATATCAGCAGCCACATCATCATCATCATCAT 192
QY 1014 GluLysSerProLeuIleAspSerTyriSerProHisGlyLeuLysProAsnThrLeuGlu 1033
Db 193 GAAAAAACCCCTTTGATTGACAGCTACAGCAGGAAGCCCTATATGCCGAACACATTGGAA 252
QY 1034 GlyAsnValThrPheAsnGluValValPheAsnTyriProThrArgProAspIleProVal 1053
Db 253 GGAATGTCACATTTGGTGAAGTTGTATCAACTATATCCACCCGCGGACATCCAGTG 312
QY 1054 LeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrLeuAlaLeuValGlySerSer 1073
Db 313 CTTCAGGACTGAGCCCTGGAGGTGAAGAAGGCCAGACGCTGCTGTGGGACAGT 372
QY 1074 GlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyriAspProLeuAlaGly 1093
Db 373 GGCTGTGGGAAGACACAGATGGTCCAGCTCTCTGGAGGGTCTTACAGCCCTTGGGAGG 432
QY 1094 SerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGlnTrpLeuArgAlaHis 1113
Db 433 AAGTGTCTGTGATGGCAAGAAATTAAGCAGTGAATGTTCACTGGCTCCGAGCACAC 492
QY 1114 LeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAla 1133
Db 493 CTGGGCAATGTCCTCCAGGAGCCCTCTCTTTGACTGCAGCATGCTCAGAACATTTGCC 552
QY 1134 TyrglyAspAsnSerArg-ValValSerHisGluGluIleMetGlnAlaAlaLysGluAl 1153
Db 553 TATGAGACAAACAGCCGGGTGTGTCTACAGGAAGAGATTTGAGGGCAGCAAGAGGAGG 512
QY 1153 aasnIleHisPheIleGluThrLeuProGluLysTyriAsnThrArgValGlyAspIly 1173
Db 613 CAACATACATGCTCTCATCGAGTCACTGCTAT-AAATATAGCACTAAACTANGAGACAA 671
QY 1173 sGlyThrGlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaAargAlaVala 1193
Db 672 AGGACCTCAGCTCTCTGGGGGCCAGAAACA-CACATTGCCATAGCTTGGCCCTGCT-AG 729
QY 1193 gGlnProHisIleLeuLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSer 1211
Db 730 ACAGACTTATAATTTGTTTGGATGAAGCCAGCTCAGCTCTTGATACAGAAAGT 784

RESULT 9
BF313560 LOCUS 795 bp mRNA linear EST 21-NOV-2000
DEFINITION 601900192F1 NTH_MGC_19 Homo sapiens cDNA clone IMAGE:4129222 5',
mRNA sequence.
ACCESSION BF313560
VERSION BF313560.1 GI:11261583
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 795)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM1023 row: h column: 23
High quality sequence stop: 674.
Location/Qualifiers
1..795
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

Qy 1255 sclythrhisglnGlnLeuLeuAlaGlnLysGlyIleTyrPheSer 1270
 Db 711 ATGGACGCATCAGAGGTGCTGGACAGAAAGG--CTCTATTTTCA 753
 RESULT 10
 BG293345 726 bp mRNA linear EST 21-FEB-2001
 LOCUS 602390738F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4502552 5',
 DEFINITION mRNA sequence.
 ACCESSION BG293345
 VERSION BG293345.1 GI:13052943
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 726)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M10371 row: 1 column: 09
 High quality sequence stop: 658.
 Location/Qualifiers
 1..726
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:450252"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 189 a 220 c 190 g 127 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3.52e-74 Length: 726
 Score: 916.00 Matches: 185
 Percent Similarity: 94.66% Conservative: 10
 Best Local Similarity: 89.81% Mismatches: 9
 Query Match: 14.16% Indels: 2
 DB: 10 Gaps: 0

US-09-672-725C-4 (1-1280) x BF313560 (1-795)
 Qy 1016 SerProLeuLeuAspSerTyrSerProHisGlyLeuLysProAsnThrLeuGluGlyAsn 1035
 Db 2 AACCCCTTTGATGACACGTACAGCGGAAGGCTTAATCCGACACATG-CAAGGAAT 60
 Qy 1036 ValThrPheAsnGluValAlaPheAsnTyrProThrArgProAspIleProValLeuGln 1055
 Db 61 GTCACATT-GGTGAAGT-GTATTCAACTATCCACCGCCGACGACATCCAGTGTTCAG 118
 Qy 1056 GlyLeuSerLeuGluValLysGlyGlnThrLeuAlaLeuValGlySerSerGlyCys 1075
 Db 119 GGACTGAGCTGGAGGTGAAGAGGCGCAGCGTGGCTCTGTGGCGACGATGGCTGT 178
 Qy 1076 GlyLysSerThrValValGlnLeuGluArgPheTyrAspProLeuAlaGlySerVal 1095
 Db 179 GGGAGAGACACAGTGGTCCAGCTCTCGAGCGGTCTACGCCCTTGGCAGGGAAGTG 238
 Qy 1096 LeuLeuAspGlyLysGluLeuLysHisLeuAsnValGlnThrPheArgAlaHisLeuGly 1115
 Db 239 CTGCTTGATGTCAGAAAGAAATAAAGCGACTGAATGCTTCCAGTCCGACACCTGGC 298
 Qy 1116 IleValSerGlnGluProIleLeuPheAspCysSerIleAlaGluAsnIleAlaTyrGly 1135
 Db 299 ATCGGTGTCAGAGGACCCATCTGTTGACGTGACGATGCTGAGACATGTCCTATGGA 358
 Qy 1136 AspAsnSerArgValValSerHisGluGluLeuMetGlnAlaAlaLysGluAlaAsnIle 1155
 Db 359 GACACACCGGGTGTGTACAGAGAGAGATCGTGAGGCGAGCAAGAGGAGGCCAACA 418
 Qy 1156 HisHisPheIleGluThrLeuProGluLysTyrAsnThrArgValGlyAspLysGlyThr 1175
 Db 419 CATGCTTTCATCGAGTCTACTGCTTAATAATATAGACATCAAGTAGAGACAAAGAACT 478
 Qy 1176 GlnLeuSerGlyGlyGlnLysGlnArgIleAlaIleAlaArgAlaLeuValArgGlnPro 1195
 Db 479 CAGCTCTCTGGTGGCCAGAACACAGCATTCGCATAGCTCGTCCCTTGTGTAGACGCT 538
 Qy 1196 HisIleLeuLeuAspGluAlaThrSerAlaLeuAspThrGluSerGluLysVal-va 1215
 Db 539 CATATTT-GCTTTGGATGAAGCCACGTCAGCTCTGGATACAGAAAGTGAAGGTGCGT 596
 Qy 1215 IGLnGluAlaLeuAspLysAlaArgGluGlyArgThrCysIleValIleAlaHisArgLe 1235
 Db 597 CCAAGAGCCCTGGACAAAGCCAGAGAAGC---GACCTGATGTGATGTGTCCACCGCTG 653
 Qy 1235 uSerThrIleGlnAsnAlaAspLeuIleValValPheGlnAsnGlyLysValLysGluHI 1255
 Db 654 -TCCACCATCCAGATGAGA--CTTATAGTGTGTTTTCAGAAATGCGAGAGTCCAGGGGC 710

QY 1147 MetGlnAlaLysGluAlaAsnIleHisPheIleGluThrLeuProGluLysTyr 1166
 Db 242 GTGAGGGCAGCAAGGAGCCAAACATCACCCCTCATCGAGACGCTGCCCAAAATAT 301
 QY 1167 AsnThrArgValGlyAspLysGlyThrGlnLeuSerGlyGlnLysGlnArgIleAla 1186
 Db 302 AACACAAGATGAGACAGGAGGACGAGCTCTCTGGGGCCAGAAAGCAGAGGATTGCC 361
 QY 1187 IleAlaArgAlaLeuValArgGlnProHisIleLeuLeuLeuAspGluAlaThrSerAla 1206
 Db 362 ATCCGCCAGCCCTCATCAGACGCTCGGCTCTACTCTGGATGAAGCCACGTCAGCT 421
 QY 1207 LeuAspThrGluSerGluLysValValGlnGluAlaLeuAspLysAlaArgGluGlyArg 1226
 Db 422 CTGGATACTAGAGTGAAGAGTTGTCCAGGAAGCAGCTGGACAAAGCCAGGAAGGCCGC 481
 QY 1227 ThrCysIleValIleAlaHisArgLeuSerThrIleGlnAsnAla-AspLeuIleValVa 1246
 Db 482 ACCTGCATTCGATCGCTACCGCTGTCCACCATCAGAACGCGCACTTGATCGTGGT 541
 QY 1246 lPheGlnAsnGlyLysValLysGluHisGlyThr-HisGlnGlnLeuLeuAlaGlnLysG 1266
 Db 542 GATTGAGACGCGAGTCAAGGACGAGCGCACCCACCCACAGCAGCTGCTGGCGCAGAAGG 601
 QY 1266 lylleTyrPheSer 1270
 Db 602 GCATCTATTCTCA 615
 RESULT 11
 BM053000
 LOCUS
 DEFINITION
 le69a07.y3 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 cDNA 5' similar to SW:MDRI_HUMAN P08183 MULTIDRUG RESISTANCE
 PROTEIN 1 ; mRNA sequence.
 ACCESSION
 BM053000
 VERSION
 BM053000.1 GI:16808894
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 587)
 AUTHORS
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemishka, I., Scaer, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Bilstain, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
 M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
 Jackson, X. and Bowers, Y.
 TITLE
 Endocrine Pancreas Consortium
 JOURNAL
 Unpublished (2000)
 COMMENT
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brownefas.harvard.edu)
 High quality sequence stop: 437.
 Location/Qualifiers
 1..587
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
 /sex="Both"
 /tissue_type="Islets of Langerhans"
 /dev_stage="Adult"
 /lab_host="DH10B"
 /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;

Site_2: Sal 1; Starting library constructed using
 Superscript Plasmid Library kit (Life Technologies). cDNA
 made by oligo-dT priming. Size-selected by column
 fractionation; average insert size 1.08 kb. Library was
 amplified once on solid support and plasmid DNA from
 library was prepared. The library DNA was normalized by
 method #4 from Bonaldo, Lennon, and Soares 1996 Genome
 Research 6:791-806; 0.5 microgram single-stranded library
 plasmid DNA was mixed with 5 micrograms PCR product
 representing library inserts and hybridized to an Ecot of
 20. Single-stranded (unhybridized) plasmids were isolated
 by hydroxyapatite chromatography and used to make this
 library."

BASE COUNT 142 a 138 c 161 g 145 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 9.6e-74 Length: 587
 Score: 910.00 Matches: 175
 Percent Similarity: 96.34% Conservative: 9
 Best Local Similarity: 91.62% Mismatches: 7
 Query Match: 14.07% Indels: 0
 DB: 10 Gaps: 0

US-09-672-725C-4 (1-1280) x BM053000 (1-587)

QY 949 MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu 968
 Db 3 ATGATGATATTTTCTCTATCTGATGTTTCCGGTTTGGAGCCTACTTGGTGCCACATAAA 62
 QY 969 PheMetAsnPheGlnAspValLeuLeuValPheSerAlaIleValPheGlyAlaMetAla 988
 Db 63 CTCATGAGCTTTGAGATGTTCTGTAGTATTTTTCAGTCTGTGTGTGGTCCCATGCC 122
 QY 989 ValGlyGlnValSerSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis 1008
 Db 123 GTGGGCAAGTCAGTTCATTTGCTCTGACTATGCCAAAGCCAAATATCAGACGCCAC 182
 QY 1009 ValIleMetIleGluLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys 1028
 Db 183 ATCATCATGATCATTTGAAAAACCCCTTTGATTCAGCTACAGCAGCAAGGCGCTAATG 242
 QY 1029 ProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArg 1048
 Db 243 CCGACACATTCGGAAGAAATGTCATTTGTTGAAGTTGTTTATTCATATCCCCACCGA 302
 QY 1049 ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysLysGlyGlnThrLeuAla 1068
 Db 303 CCGGACATCCCATGCTTCAGGGACTGAGCCTGGAGGTGAAGAAGGCCAGACGCTGGCT 362
 QY 1069 LeuValGlySerSerGlyCysGlyLysSerThrValValGlnLeuLeuGluArgPheTyr 1088
 Db 363 CTGGTGGGCAGCAGTGGCTGTGGGAAGAGCAGTGGTCCAGCTCCTGGAGCGGTTCATC 422
 QY 1089 AspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGln 1108
 Db 423 GACCCCTTGGCAGGAAAGTCTGCTTCATGGCAAGAAATAAAGCGACTGAATGTTCTAG 482
 QY 1109 TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle 1128
 Db 483 TGGTCTCCGAGCACACCTGGCATCTGTGCCAGGAGGCCATCTCTTTGACTGTCAGCAT 542
 QY 1129 AlaGluAsnIleAlaTyrGlyAspAsnSerArg 1139
 Db 543 GCTGAGAACATTTGCTATGGAGACAACAGCCGN 575

RESULT 12
 BM080311
 LOCUS
 DEFINITION
 H3052B06-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 H3052B06 5', mRNA sequence.
 ACCESSION
 BM080311
 VERSION
 BM080311.1 GI:12562879

FEATURES
 source

BG080311 628 bp mRNA linear EST 26-JAN-2001
 H3052B06-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 H3052B06 5', mRNA sequence.
 BG080311
 BG080311.1 GI:12562879

Score: 889.00 Matches: 171
Percent Similarity: 96.26% Conservatives: 9
Best Local Similarity: 91.44% Mismatches: 7
Query Match: 13.75% Indels: 0
DB: 10 Gaps: 0

US-09-672-725c-4 (1-1280) x BF964218 (1-563)

QY	949	MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu	968
DB	3	ATGATGATATTTTCCCTATGCTGGATGTTTCCGGTTTGGAGCCTACTTGTGGCACAATAA	62
QY	969	PhMetAsnPheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAla	988
DB	63	CTCATGAGCTTTCAGATGTTCTGTAGTATTTTCAGCTGTGCTTGGTCCATGGCC	122
QY	989	ValGlyGlnValSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis	1008
DB	123	GTGGGCAAGTCAGTTCATTTGCTGCTGACTATGCCAAAGCCAAATATATCAGCAGCCAC	182
QY	1009	ValIleMetIleGlyLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys	1028
DB	183	ATCATCATGATCATTTGAAAAAACCCCTTTGATTGACAGCTACAGCAGGAGCCCTAATG	242
QY	1029	ProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArg	1048
DB	243	CCGACACATTTGGAGGAATGTCACATTTTGGTGAAGTGTATTAACATATCCACCCGA	302
QY	1049	ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAla	1068
DB	303	CCGACATCCAGTCTTTCAGGGACTGAGCTGGAGGTGAAGAAGGCCAGCGCTGGCT	362
QY	1069	LeuValGlySerGlyCysGlyLysSerThrValValGlnLeuGluArgPheTyr	1088
DB	363	CTGGTGGCAGCAGTGGCTGTGGAGAGACACAGTGGTCCAGCTCTGGAGCGGTCTAC	422
QY	1089	AspProLeuAlaGlySerValLeuIleAspGlyLysGluLeuLysHisLeuAsnValGln	1108
DB	423	GACCCCTTGGCAGGAAAGTGTCTGTGATGGCAAGAAATAAAGGACTGAATGTTACG	482
QY	1109	TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle	1128
DB	483	TGGCTCCGAGCACACCTTGGGAGTGTGTCCAGAGCCCATCTGTTGACTCCAGCATT	542
QY	1129	AlaGluAsnIleAlaTyrGly	1135
DB	543	GCTGAGAACATGGCTATGGA	563

RESULT 15
BM052878 563 bp mRNA linear EST 08-NOV-2001
LOCUS 167e03.y3 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION cDNA 5' similar to SW:MDR1_HUMAN P08183 MULTIDRUG RESISTANCE
PROTEIN 1 ; mRNA sequence.

ACCESSION BM052878
VERSION BM052878.1 GI:16808672
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 563)
AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarelis, V., Williams, T., Jackson, Y., and Bowers, Y.

TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)
High quality sequence stop: 430.
Location/Qualifiers
1. 563

FEATURES source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript plasmid library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. the library DNA was normalized by method #4 from Bonaudo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."
BASE COUNT 136 a 132 c 153 g 142 t
ORIGIN

Alignment Scores:
Pred. No.: 7,84e-72 Length: 563
Score: 889.00 Matches: 171
Percent Similarity: 96.26% Conservatives: 9
Best Local Similarity: 91.44% Mismatches: 7
Query Match: 13.75% Indels: 0
DB: 10 Gaps: 0

US-09-672-725c-4 (1-1280) x BM052878 (1-563)

QY	949	MetMetTyrPheSerTyrAlaGlyCysPheArgPheGlyAlaTyrLeuValAlaAsnGlu	968
DB	3	ATGATGATATTTTCCCTATGCTGGATGTTTCCGGTTTGGAGCCTACTTGTGGCACAATAA	62
QY	969	PhMetAsnPheGlnAspValLeuValPheSerAlaIleValPheGlyAlaMetAla	988
DB	63	CTCATGAGCTTTCAGATGTTCTGTAGTATTTTCAGCTGTGCTTGGTCCATGGCC	122
QY	989	ValGlyGlnValSerPheAlaProAspTyrAlaLysAlaLysValSerAlaAlaHis	1008
DB	123	GTGGGCAAGTCAGTTCATTTGCTGCTGACTATGCCAAAGCCAAATATATCAGCAGCCAC	182
QY	1009	ValIleMetIleGlyLysSerProLeuIleAspSerTyrSerProHisGlyLeuLys	1028
DB	183	ATCATCATGATCATTTGAAAAAACCCCTTTGATTGACAGCTACAGCAGGAGCCCTAATG	242
QY	1029	ProAsnThrLeuGluGlyAsnValThrPheAsnGluValValPheAsnTyrProThrArg	1048
DB	243	CCGACACATTTGGAGGAATGTCACATTTTGGTGAAGTGTATTAACATATCCACCCGA	302
QY	1049	ProAspIleProValLeuGlnGlyLeuSerLeuGluValLysGlyGlnThrLeuAla	1068
DB	303	CCGACATCCAGTCTTTCAGGGACTGAGCTGGAGGTGAAGAAGGCCAGCGCTGGCT	362
QY	1069	LeuValGlySerGlyCysGlyLysSerThrValValGlnLeuGluArgPheTyr	1088
DB	363	CTGGTGGCAGCAGTGGCTGTGGAGAGACACAGTGGTCCAGCTCTGGAGCGGTCTAC	422

```

QY 1089 AspProLeuAlaGlySerValLeuIleAspGlyLysGluIleLysHisLeuAsnValGln 1108
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 GACCCCTGGCAGGGAAGTGCTGCTGATGGCAAGAAATAAAGCGACTGAATGTTTCA 482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1109 TrpLeuArgAlaHisLeuGlyIleValSerGlnGluProIleLeuPheAspCysSerIle 1128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 483 TGGCTCCGAGCACACCTGGGGCATCGTGCCAGGAGCCCATCCCTGTTGACTGCAGCATT 542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1129 AlaGluAsnIleAlaTyrGly 1135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 GCTGAGAACATTGCCTATGGA 563
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: November 7, 2002, 07:28:27
Job time : 2256.22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 17:45:54 ; Search time 9.04753 Seconds
(without alignments)
4517.824 Million cell updates/sec

Title: US-09-672-725C-6

Perfect score: 1836

Sequence:

1 TSLDTESEAVVQVALDKAR.....TVVSLTQEQKEHMYAQSLQ 368

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1836	100.0	368 22 AAB81069	Dog P-glycoprotein
2	1836	100.0	368 22 AAE00305	Dog P-glycoprotein
3	1808	98.5	1281 22 AAE00303	Dog (PGP) P-glycop
4	1808	98.5	1281 22 AAE00308	Dog P-glycoprotein
5	1808	98.5	1281 22 AAE00309	Dog P-glycoprotein
6	1808	98.5	1281 22 AAE00310	Dog P-glycoprotein
7	1800	98.0	1280 22 AAB81068	Dog P-glycoprotein
8	1800	98.0	1280 22 AAE00304	Dog P-glycoprotein
9	1678	91.4	1280 22 AAB81064	Cynomolgous monke
10	1678	91.4	1283 22 AAB81065	Cynomolgous monke
11	1672	91.1	1279 19 AAW80294	Mutated human P-gi

12	1672	91.1	1279	22	AAB81067	Human P-glycoprote
13	1672	91.1	1279	22	AAE00307	Human P-glycoprote
14	1672	91.1	1280	8	AAAP70452	Sequence encoded b
15	1672	91.1	1280	15	AAAG63624	Human P-glycoprote
16	1672	91.1	1280	18	AAW44073	Human multidrug re
17	1672	91.1	1280	21	AAV58186	Human P-glycoprote
18	1672	91.1	1280	21	AAV58187	Human P-glycoprote
19	1672	91.1	1280	22	AAB81066	Human G185V mutat
20	1672	91.1	1280	22	AAB81066	Human P-glycoprote
21	1672	91.1	1280	22	AAB81959	Human MDR1. Homo
22	1672	91.1	1280	22	AAE00306	Human P-glycoprote
23	1669	90.9	1280	11	AAE04868	Protein encoded by
24	1665	90.7	1280	14	AAE44297	Sequence encoded b
25	1662	90.5	1280	22	AAU04347	Human multidrug re
26	1644	89.5	1280	19	AAW48997	Wild-type human p
27	1644	89.5	1280	19	AAW48998	N-terminal single
28	1644	89.5	1280	19	AAW48999	C-terminal single
29	1644	89.5	1280	19	AAW49000	Human P glycoprote
30	1575	85.8	1276	21	AAV58189	Murine multidrug r
31	1575	85.8	1272	21	AAV70596	Rat multidrug resi
32	1426.5	77.7	1275	21	AAV78879	Rat multidrug resi
33	1426.5	77.7	1275	22	AAE60409	Rat mdr1b2 (multis
34	1401.5	76.3	1276	14	AAE35199	Rat mdr1b2 multidr
35	1401.5	76.3	1276	21	AAV58188	Mouse multidrug re
36	664.5	36.2	1275	20	AAW82594	Murine multidrug r
37	660	35.9	1313	22	ABB63904	H. contortus PGP-A
38	598	32.6	1279	22	ABB60234	Drosophila melanog
39	597.5	32.5	1302	22	ABB64919	Drosophila melanog
40	591	32.2	1334	20	AAV02630	Protein encoded by
41	588.5	32.1	1349	20	AAV16434	Multiple drug resi
42	583	31.8	1320	22	ABB64922	Drosophila melanog
43	564	30.7	1254	21	AAV92173	A. thaliana ATPAC,
44	558	30.4	1205	21	AAG39103	Arabidopsis thalia
45	558	30.4	1243	21	AAG39102	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAB81069

ID AAB81069 standard; Protein; 368 AA.

XX AC AAB81069;

XX DT 25-JUN-2001 (first entry)

XX DE Dog P-glycoprotein SEQ ID 8.

XX KW Cynomolgous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL;

XX OS efflux pump; dog.

XX OS Canis familiaris.

XX PN WO200123565-A1.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26592.

XX PR 28-SEP-1999; 99US-0156921.

XX PR 12-OCT-1999; 99US-0158818.

XX (GENT-) GENTEST CORP.

XX Stocker PJ, Steimel-Crespi DT, Crespi CL;

XX WPI; 2001-316136/33.

XX Novel isolated nucleic acid encoding cynomolgous monkey P-glycoprotein

XX (PGP) and homologous PGP polypeptides are useful for predicting

XX bioavailability of compound and increasing PGP transporter activity in

XX cell -

XX

XX

XX PS Claim 8; Page 77-77; 84pp; English.

XX CC This invention relates to a polynucleotide sequence encoding a

CC cynomolgus monkey P-glycoprotein (PGP), and an allelic variant of the

CC PGP protein. PGP, also known as multidrug transporter, MDR1 is a member

CC of the ABC transporter superfamily. The enzyme serves as an efflux pump

CC exporting small molecules across the cell membrane. The invention

CC includes a cynomolgus monkey (Macaca fascicularis) PGP coding sequence

CC and protein, and also that of an allelic variant. The PGP polynucleotide

CC sequence is useful for increasing PGP transporter activity in a cell.

CC Antisense sequences of the cDNA are useful for inhibiting PGP transport

CC activity in a mammalian cell. They may also be used for increasing the

CC bioavailability of a drug. The present sequence represents

CC P-glycoprotein from a dog.

XX SQ Sequence 368 AA;

Query Match 100.0%; Score 1836; DB 22; Length 368;

Best Local Similarity 100.0%; Pred. No. 5.3e-181;

Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 60

Db 1 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 60

Qy 61 EKGIFYKLVMTQTRNGNEIDLENATGESKESDALEMSPKDSGLIKRRSTRSIHAPOG 120

Db 61 EKGIFYKLVMTQTRNGNEIDLENATGESKESDALEMSPKDSGLIKRRSTRSIHAPOG 120

Qy 121 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFVVGIFCAIINGGLQPAFSIFSRIG 180

Db 121 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFVVGIFCAIINGGLQPAFSIFSRIG 180

Qy 181 IFTRDEDPETKROISNMFSVLFLVGLGIISFTFFLQGTFGKAGEILTCLRMYMFRSML 240

Db 181 IFTRDEDPETKROISNMFSVLFLVGLGIISFTFFLQGTFGKAGEILTCLRMYMFRSML 240

Qy 241 RQDVSWFDDLKNTTGALTTRLANDAAQVKAIGSLAVITONIANLGTGIIISLIYGWQL 300

Db 241 RQDVSWFDDLKNTTGALTTRLANDAAQVKAIGSLAVITONIANLGTGIIISLIYGWQL 300

Qy 301 TLLLLAIVPIITAAGVWEKMLSGQALKDKKEGAGKIATEAIENFRTVVSILTQEQKFE 360

Db 301 TLLLLAIVPIITAAGVWEKMLSGQALKDKKEGAGKIATEAIENFRTVVSILTQEQKFE 360

Qy 361 HMYAQSLQ 368

Db 361 HMYAQSLQ 368

RESULT 2

AAE00305

ID AAE00305 standard; Protein; 368 AA.

XX AC AAE00305;

DT 13-JUN-2001 (first entry)

XX DE Dog P-glycoprotein (PGP) #3.

XX KW Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;

XX KW drug bioavailability; transgenic animal; genetic model.

XX OS Canis familiaris.

XX PN WO200123540-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26767.

XX PF 28-SEP-1999; 99US-0156510.

XX PF

XX PA (GENT-) GENTEST CORP.

XX PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;

XX DR WPI; 2001-235373/24.

XX DR N-PSDB; AAD03490.

XX PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful

PT for determining the bioavailability of drugs and for screening for dog

PT PGP inhibitors -

XX PS Disclosure; Page 77; 11pp; English.

XX CC The invention relates to dog P-glycoprotein (PGP) also referred

CC as multidrug transporter (MDR1) and nucleic acids encoding them.

CC The invention also includes fragments and biologically functional

CC variants of dog P-glycoprotein. PGP and their nucleic acids are

CC useful for determining the bioavailability of drugs and for

CC screening PGP inhibitors. They are useful for the diagnosis and

CC treatment of conditions characterised by PGP activity, by

CC reducing or increasing PGP activity in a cell. PGP nucleic acids

CC are used as oligonucleotide probes. Complements of PGP nucleic

CC acids are useful as antisense oligonucleotides, to induce a PGP

CC 'knockout' phenotype. They are used to prepare a non-human

CC transgenic animal, which are valuable as genetic models for

CC human diseases.

CC The present sequence is dog P-glycoprotein (PGP). The

CC PGP enzyme functions as an efflux pump exporting small molecules

CC across the cell membrane. This enzyme is a member of the ABC

CC transporter family.

XX SQ Sequence 368 AA;

Query Match 100.0%; Score 1836; DB 22; Length 368;

Best Local Similarity 100.0%; Pred. No. 5.3e-181;

Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 60

Db 1 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 60

Qy 61 EKGIFYKLVMTQTRNGNEIDLENATGESKESDALEMSPKDSGLIKRRSTRSIHAPOG 120

Db 61 EKGIFYKLVMTQTRNGNEIDLENATGESKESDALEMSPKDSGLIKRRSTRSIHAPOG 120

Qy 121 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFVVGIFCAIINGGLQPAFSIFSRIG 180

Db 121 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFVVGIFCAIINGGLQPAFSIFSRIG 180

Qy 181 IFTRDEDPETKROISNMFSVLFLVGLGIISFTFFLQGTFGKAGEILTCLRMYMFRSML 240

Db 181 IFTRDEDPETKROISNMFSVLFLVGLGIISFTFFLQGTFGKAGEILTCLRMYMFRSML 240

Qy 241 RQDVSWFDDLKNTTGALTTRLANDAAQVKAIGSLAVITONIANLGTGIIISLIYGWQL 300

Db 241 RQDVSWFDDLKNTTGALTTRLANDAAQVKAIGSLAVITONIANLGTGIIISLIYGWQL 300

Qy 301 TLLLLAIVPIITAAGVWEKMLSGQALKDKKEGAGKIATEAIENFRTVVSILTQEQKFE 360

Db 301 TLLLLAIVPIITAAGVWEKMLSGQALKDKKEGAGKIATEAIENFRTVVSILTQEQKFE 360

Qy 361 HMYAQSLQ 368

Db 361 HMYAQSLQ 368

RESULT 3

AAE00303

ID AAE00303 standard; Protein; 1281 AA.

XX AC AAE00303;

XX AC

DT 13-JUN-2001 (first entry)
 XX
 DE Dog (PGP) P-glycoprotein (genotype C) #1.
 XX
 KW Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
 KW drug bioavailability; transgenic animal; genetic model.
 XX
 OS Canis familiaris.
 XX
 PN WO200123540-A2.
 XX
 PD 05-APR-2001.
 XX
 XX 28-SEP-2000; 2000WO-US26767.
 PF
 XX 28-SEP-1999; 99US-0156510.
 PR
 XX (GENT-) GENTEST CORP.
 PA
 XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 PI WPI; 2001-235373/24.
 XX N-PSDB; AAD03488.
 DR
 XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors -
 XX
 XX Claim 17; Page 64-66; 111pp; English.
 PS
 XX The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (PGP) also referred
 CC as genotype C protein. The PGP enzyme functions as an efflux
 CC pump exporting small molecules across the cell membrane. This
 CC enzyme is a member of the ABC transporter family.
 XX
 XX Sequence 1281 AA;
 SQ
 Query Match 98.5%; Score 1808; DB 22; Length 1281;
 Best Local Similarity 98.6%; Pred. No. 2.5e-177;
 Matches 363; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TSALDTESAVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVIVEKGNHDELMK 60
 DB 560 TSALDTESAVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVIVEKGNHDELMK 619
 QY 61 EKGIFYKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRSTRSIHAPOQ 120
 DB 620 EKGIFYKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRSTRSIHAPOQ 679
 QY 121 QDRKLGTKEDLNENPPVSFWRILKLNSTWPFYFVVGIFCAIINGGLQPAFSIIFSLIG 180
 DB 680 QDRKLGTKEDLNENPPVSFWRILKLNSTWPFYFVVGIFCAIINGGLQPAFSIIFSLIG 739
 QY 181 IFRDREDPTKQISNMESVFLVLGIIISFTFFLQGGTFPGKAGEILTTLRLRYMVFRSML 240
 DB 740 IFRDREDPTKQISNMESVFLVLGIIISFTFFLQGGTFPGKAGEILTTLRLRYMVFRSML 799
 QY 241 RQDVSWFDDLNKNTTGALTTRLANDAAQVKGAIGSLRLAVITQNIANLGTGIIISLIYQWL 300
 DB 800 RQDVSWFDDPNKTTGALTTRLANDAAQVKGAIGSLRLAVITQNIANLGTGIIISLIYQWL 859

QY 301 TLLLLAIPIITAIAGVWEMKMLSGALKKKEGAGKATEAIENTRTVSVLTQOKFE 360
 DB 860 TLLLLAIPIITAIAGVWEMKMLSGALKKKEGAGKATEAIENTRTVSVLTQOKFE 919
 QY 361 HMYAQSLQ 368
 DB 920 YMYAQSLQ 927

RESULT 4
 AAE00308
 ID AAE00308 standard; Protein; 1281 AA.
 XX
 AC AAE00308;
 XX
 DT 13-JUN-2001 (first entry)
 XX
 DE Dog P-glycoprotein (PGP) allelic variant (Genotype A).
 XX
 KW Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
 KW MDR1; drug bioavailability; transgenic animal; genetic model.
 XX
 OS Canis familiaris.
 XX
 FH Key Location/Qualifiers.
 FT Misc-difference 197
 FT /note= "His of GenotypeC substituted by Gln"
 XX
 PN WO200123540-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26767.
 XX
 PR 28-SEP-1999; 99US-0156510.
 XX
 XX (GENT-) GENTEST CORP.
 PA
 XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 PI WPI; 2001-235373/24.
 XX N-PSDB; AAD03504.
 DR
 XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors -
 XX
 XX Claim 17; Page 91-93; 111pp; English.
 PS
 XX The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
 CC This sequence is also referred as Genotype A protein. The PGP
 CC enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.
 XX
 SQ Sequence 1281 AA;
 Query Match 98.5%; Score 1808; DB 22; Length 1281;
 Best Local Similarity 98.6%; Pred. No. 2.5e-177;

FT Misc-difference 329 /note= "His of GenotypeC substituted by Gln"
FT /note= "Ser of GenotypeC substituted by Thr"
FT Misc-difference 1148
FT /note= "Met of GenotypeC substituted by Val"
XX
PN WO200123540-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26767.
XX
PR 28-SEP-1999; 99US-0156510.
XX
PA (GENT-) GENTEST CORP.
XX
PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX
DR WPI; 2001-235373/24.
DR N-PSDB; AAD03506.
XX
PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT for determining the bioavailability of drugs and for screening for dog
PT PGP inhibitors -
XX
PS Claim 17; Page 108-110; 11lpp; English.
XX
CC The invention relates to dog P-glycoprotein (PGP) also referred
CC as multidrug transporter (MDR1) and nucleic acids encoding them.
CC The invention also includes fragments and biologically functional
CC variants of dog P-glycoprotein. PGP and their nucleic acids are
CC useful for determining the bioavailability of drugs and for
CC screening PGP inhibitors. They are useful for the diagnosis and
CC treatment of conditions characterised by PGP activity, by
CC reducing or increasing PGP activity in a cell. PGP nucleic acids
CC are used as oligonucleotide probes. Complements of PGP nucleic
CC acids are useful as antisense oligonucleotides, to induce a PGP
CC 'knockout' phenotype. They are used to prepare a non-human
CC transgenic animal, which are valuable as genetic models for
CC human diseases.
CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
CC This sequence is also referred as Genotype D protein. The PGP
CC enzyme functions as an efflux pump exporting small molecules
CC across the cell membrane. This enzyme is a member of the ABC
CC transporter family.
XX
SQ Sequence 1281 AA;

Query Match 98.5%; Score 1808; DB 22; Length 1281;
Best Local Similarity 98.6%; Pred. No. 2.5e-177;
Matches 363; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 TSALDTESEAVQVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 60
DB 560 TSALDTESEAVQVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 619
QY 61 EKGIVFKLVTMOTRGNIELEENATGESKESDALEMSPKDSGLIKRRSTRSHAPQG 120
DB 620 EKGIVFKLVTMOTRGNIELEENATGESKESDALEMSPKDSGLIKRRSTRSHAPQG 679
QY 121 QDRKLGTKEDLNENPVPSFWRILKLNSTWEPYFVVGIFCAINGLQPAFSIIFRSIIG 180
DB 680 QDRKLGTKEDLNENPVPSFWRILKLNSTWEPYFVVGIFCAINGLQPAFSIIFRSIIG 739
QY 181 IFTREDPEPKRQISNMFSVLFLVGLGISFTFFLQGGTFFGKAGBILTKRLRYWVFRSML 240
DB 740 IFTREDPEPKRQSNKFSVLFLVGLGISFTFFLQGGTFFGKAGBILTKRLRYWVFRSML 799
QY 241 RQDVSWFDDLNKTGALTFRRLANDAQAQVKGAGISRLAVITONIANLGTGIIISLIYGWQL 300
DB 800 RQDVSWFDDPKNTGALTFRRLANDAQAQVKGAGISRLAVITONIANLGTGIIISLIYGWQL 859
QY 301 TLLLLAIVPIIAIAGVWEMKLSGOALKDKKEGAGKIATEIENFTVVSILTREQKFE 360

DB 860 TLLLLAIVPIIAIAGVWEMKLSGOALKDKKEGAGKIATEIENFTVVSILTREQKFE 919
QY 361 HMYAQLSQ 368
DB 920 YMYAQLSQ 927
RESULT 7
AAB81068
ID AAB81068 standard; Protein; 1280 AA.
XX
AC AAB81068;
XX
DT 25-JUN-2001 (first entry)
DE Dog P-glycoprotein SEQ ID 7.
XX
KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW efflux pump; dog.
XX
OS Canis familiaris.
XX
PN WO200123565-A1.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26592.
XX
PR 28-SEP-1999; 99US-0156921.
PR 12-OCT-1999; 99US-0158818.
XX
PA (GENT-) GENTEST CORP.
XX
PI Stocker PJ, Steimel-Crespi DT, Crespi CL;
XX WPI; 2001-316136/33.
XX
PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell -
XX
PS Claim 8; Page 73-76; 84pp; English.
XX
CC This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDR1 is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents
CC P-glycoprotein from a dog.
XX
SQ Sequence 1280 AA;

Query Match 98.0%; Score 1800; DB 22; Length 1280;
Best Local Similarity 98.4%; Pred. No. 1.7e-176;
Matches 362; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 TSALDTESEAVQVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 60
DB 559 TSALDTESEAVQVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 618
QY 61 EKGIVFKLVTMOTRGNIELEENATGESKESDALEMSPKDSGLIKRRSTRSHAPQG 120
DB 619 EKGIVFKLVTMOTRGNIELEENATGESKESDALEMSPKDSGLIKRRSTRSHAPQG 678
QY 121 QDRKLGTKEDLNENPVPSFWRILKLNSTWEPYFVVGIFCAINGLQPAFSIIFRSIIG 180

||||| 679 QDRKLTGKEDLNENPVSVFWRILKLNSTENPFFVVGIFCAIINGGLQPAFSLIISRIIG 738
181 IFTREDEPETHKQISNMFSVLFLVLGIISFTIFFLQGFTEGKAGEILTKRLRYMVFRSML 240
739 IFTREDEPETHKQNSNMFSVLFLVLGIISFTIFFLQGFTEGKAGEILTKRLRYMVFRSML 798
241 RQDVSNFDDPKNTTGALTTRLANDAAQVKGAGSLAVITONTANLGTGIIISLIYGWOL 300
799 RQDVSNFDDPKNTTGALTTRLANDAAQVKGAGSLAVITONTANLGTGIIISLIYGWOL 858
301 TLLLLAIVPIIATAGVVEKMLSGQALKDKKELEGAGKIATEAIENFRVWSLTQEQKFE 360
859 TLLLLAIVPIIATAGVVEKMLSGQALKDKKELEGAGKIATEAIENFRVWSLTQEQKFE 918
361 HMYAQSILQ 368
919 YMYAQSILQ 926
RESULT 8
AAE00304
ID AAE00304 standard; Protein: 1280 AA.
XX AC AAE00304;
XX DT 13-JUN-2001 (first entry)
XX DE Dog P-glycoprotein (PGP) #2.
XX KW Dog; P-glycoprotein; PGP; multidrug transporter; MDRL;
XX KW drug bioavailability; transgenic animal; genetic model.
XX OS Canis familiaris.
XX PN WO200123540-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26767.
XX PR 28-SEP-1999; 99US-0156510.
XX PA (GENT-) GENTEST CORP.
XX PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX WPI; 2001-235373/24.
XX DR N-PSDB; AAD03489.
XX PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX PT for determining the bioavailability of drugs and for screening for dog
XX PT PGP inhibitors -
XX PS Example 2; Page 72-75; 11lpp; English.
XX CC The invention relates to dog P-glycoprotein (PGP) also referred
XX CC as multidrug transporter (MDRL) and nucleic acids encoding them.
XX CC The invention also includes fragments and biologically functional
XX CC variants of dog P-glycoprotein. PGP and their nucleic acids are
XX CC useful for determining the bioavailability of drugs and for
XX CC screening PGP inhibitors. They are useful for the diagnosis and
XX CC treatment of conditions characterised by PGP activity, by
XX CC reducing or increasing PGP activity in a cell. PGP nucleic acids
XX CC are used as oligonucleotide probes. Complements of PGP nucleic
XX CC acids are useful as antisense oligonucleotides, to induce a PGP
XX CC 'knockout' phenotype. They are used to prepare a non-human
XX CC transgenic animal, which are valuable as genetic models for
XX CC human diseases.
XX CC The present sequence is dog P-glycoprotein (PGP). The
XX CC PGP enzyme functions as an efflux pump exporting small molecules
XX CC across the cell membrane. This enzyme is a member of the ABC
XX CC transporter family.

XX SQ Sequence 1280 AA;
Query Match 98.0%; Score 1800; DB 22; Length 1280;
Best Local Similarity 98.4%; Pred. No. 1.7e-176;
Matches 362; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGDDGVIVKGNHDELMK 60
DB 559 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGDDGVIVKGNHDELMK 618
QY 61 EKGIVFKLVMTQTRGNEIDLENATGESKESDALEMSKDSGLIKRRSTRRSIHAPOG 120
DB 619 EKGIVFKLVMTQTRGNEIDLENATGESKESDALEMSKDSGLIKRRSTRRSIHAPOG 678
QY 121 QDRKLTGKEDLNENPVSVFWRILKLNSTENPFFVVGIFCAIINGGLQPAFSLIISRIIG 180
DB 679 QDRKLTGKEDLNENPVSVFWRILKLNSTENPFFVVGIFCAIINGGLQPAFSLIISRIIG 738
QY 181 IFTREDEPETHKQISNMFSVLFLVLGIISFTIFFLQGFTEGKAGEILTKRLRYMVFRSML 240
DB 739 IFTREDEPETHKQNSNMFSVLFLVLGIISFTIFFLQGFTEGKAGEILTKRLRYMVFRSML 798
QY 241 RQDVSNFDDPKNTTGALTTRLANDAAQVKGAGSLAVITONTANLGTGIIISLIYGWOL 300
DB 799 RQDVSNFDDPKNTTGALTTRLANDAAQVKGAGSLAVITONTANLGTGIIISLIYGWOL 858
QY 301 TLLLLAIVPIIATAGVVEKMLSGQALKDKKELEGAGKIATEAIENFRVWSLTQEQKFE 360
DB 859 TLLLLAIVPIIATAGVVEKMLSGQALKDKKELEGAGKIATEAIENFRVWSLTQEQKFE 918
QY 361 HMYAQSILQ 368
DB 919 YMYAQSILQ 926
RESULT 9
AAB81064
ID AAB81064 standard; Protein: 1280 AA.
XX AC AAB81064;
XX DT 25-JUN-2001 (first entry)
XX DE Cynomologous monkey P-glycoprotein variant 1.
XX DE Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL;
XX KW efflux pump.
XX OS Macaca fascicularis.
XX PN WO200123565-A1.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26592.
XX PR 28-SEP-1999; 99US-0156921.
XX PR 12-OCT-1999; 99US-0158818.
XX PA (GENT-) GENTEST CORP.
XX PI Stocker PJ, Steimel-Crespi DT, Crespi CL;
XX WPI; 2001-316136/33.
XX DR N-PSDB; AAF86127.
XX PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
XX PT (PGP) and homologous PGP polypeptides are useful for predicting
XX PT bioavailability of compound and increasing PGP transporter activity in
XX PT cell -
XX PS Claim 9; Page 57-59; 84pp; English.

XX This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDR1 is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents the
CC cynomologous monkey P-glycoprotein.
XX
SQ Sequence 1280 AA;

Query Match 91.4%; Score 1678; DB 22; Length 1280;
Best Local Similarity 90.5%; Pred. No. 6.7e-164;
Matches 333; Conservative 18; Mismatches 17; Indels 0; Gaps 0;

QY 1 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 60
Db 558 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 617
QY 61 EKGIFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGSLIKRSTRRSIHAPQG 120
Db 618 EKGIFKLVMTQTAGNEILENAADESKSEIDTLEMSHSDSGSLIKRSTRRSVRGSG 677
QY 121 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFVVGIFCAIINGGLQPAFISFIIRIG 180
Db 678 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFVVGIFCAIINGGLQPAFISFIIRIG 737
QY 181 IFTREDPETKRQISNMFSLVFLVGIISFTIFFLQGTFFGKAGEILTKRLRYMVFMSML 240
Db 738 IFTRNDAETKRONSLFSLFLVLGIVSFITFFLQGTFFGKAGEILTKRLRYMVFMSML 797
QY 241 RQDVSWFDDKNTGTALTRLANDAAQVKAIGSRSLAVITQNTIANLGTGIIISLIYQWL 300
Db 798 RQDVSWFDDKNTGTALTRLANDAAQVKAIGSRSLAVITQNTIANLGTGIIISLIYQWL 857
QY 301 TLLLLAIVPIIATAGVVEKMLSGQALKDKKEGAGKIATEAIEENFTVVSITQEQKFE 360
Db 858 TLLLLAIVPIIATAGVVEKMLSGQALKDKKEGAGKIATEAIEENFTVVSITQEQKFE 917
QY 361 HMYAQSLQ 368
Db 918 HMYDQSLQ 925

RESULT 10
AAB81065
ID AAB81065 standard; Protein; 1283 AA.
XX
AC AAB81065;
XX
DT 25-JUN-2001 (first entry)
XX
DE Cynomologous monkey P-glycoprotein variant 2.
XX
KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW efflux pump.
XX
OS Macaca fascicularis.
XX
FH Key Location/Qualifiers
FT Misc-difference 93-95
FT /note= "An additional 3 amino acids are present compared
FT to PGP variant AAB81064"
XX
PN WO200123565-A1.
XX
PD 05-APR-2001.
XX

PF 28-SEP-2000; 2000WO-US26592.
XX
PR 28-SEP-1999; 99US-0156921.
PR 12-OCT-1999; 99US-0158818.
XX
PA (GENT-) GENTEST CORP.
XX
PI Stocker PJ, Steimel-Crespi DT, Crespi CL;
XX WPI; 2001-316136/33.
DR N-PSDB; AAF86128.
XX
PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell -
XX
PS Claim 9; Page 65-68; 84pp; English.
XX
CC This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDR1 is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents the
CC cynomologous monkey P-glycoprotein variant 2. The protein has an
CC additional 3 amino acids when compared to PGP variant 1 (AAB81065).
XX
SQ Sequence 1283 AA;

Query Match 91.4%; Score 1678; DB 22; Length 1283;
Best Local Similarity 90.5%; Pred. No. 6.8e-164;
Matches 333; Conservative 18; Mismatches 17; Indels 0; Gaps 0;

QY 1 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 60
Db 561 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 620
QY 61 EKGIFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGSLIKRSTRRSIHAPQG 120
Db 621 EKGIFKLVMTQTAGNEILENAADESKSEIDTLEMSHSDSGSLIKRSTRRSVRGSG 680
QY 121 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFVVGIFCAIINGGLQPAFISFIIRIG 180
Db 681 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFVVGIFCAIINGGLQPAFISFIIRIG 740
QY 181 IFTREDPETKRQISNMFSLVFLVGIISFTIFFLQGTFFGKAGEILTKRLRYMVFMSML 240
Db 741 IFTRNDAETKRONSLFSLFLVLGIVSFITFFLQGTFFGKAGEILTKRLRYMVFMSML 800
QY 241 RQDVSWFDDKNTGTALTRLANDAAQVKAIGSRSLAVITQNTIANLGTGIIISLIYQWL 300
Db 801 RQDVSWFDDKNTGTALTRLANDAAQVKAIGSRSLAVITQNTIANLGTGIIISLIYQWL 860
QY 301 TLLLLAIVPIIATAGVVEKMLSGQALKDKKEGAGKIATEAIEENFTVVSITQEQKFE 360
Db 861 TLLLLAIVPIIATAGVVEKMLSGQALKDKKEGAGKIATEAIEENFTVVSITQEQKFE 920
QY 361 HMYAQSLQ 368
Db 921 HMYDQSLQ 928

RESULT 11
AAB80294
ID AAB80294 standard; Protein; 1279 AA.
XX
AC AAB80294;

XX 08-JAN-1999 (first entry)
DT Mutated human P-glycoprotein designated mdr-delta-F335/336.
XX
DE
XX mdr1-delta-F335; cyclosporin-resistant deletion mutant; human;
KW P-glycoprotein; cyclosporin; drug efflux;
KW modulator-resistant multidrug resistance; cancer therapy; ss.
XX
XX Homo sapiens.
OS
XX US5830697-A.
PN
XX 03-NOV-1998.
PD
XX 21-JAN-1997; 97US-0784649.
PF
XX 21-JAN-1997; 97US-0784649.
PR
XX (STRD) UNIV LELAND STANFORD JUNIOR.
PA
XX Chen G, Sikic BI;
PI WPI; 1998-609244/51.
XX
DR DNA encoding mutant P-glycoprotein polypeptide - useful for
XX producing recombinant polypeptide
PT
XX Claim 2; Columns 19-26; 17pp; English.
PS
XX The present sequence represents mdr1-delta-F335/336, a
XX cyclosporin-resistant deletion mutant of human P-glycoprotein lacking
CC Phe335, Phe 336 or both. The recombinant protein can be used to produce
CC antibodies. It can also be used in drug screening assays to study the
CC effects of modulators such as cyclosporin on P-glycoprotein-mediated
CC drug efflux. The nucleic acid can be used to transfer modulator-resistant
CC multidrug resistance into transfected cells, e.g. haematopoietic stem
CC cells for use during cancer therapy.
XX
XX Sequence 1279 AA;
SQ
Query Match 91.1%; Score 1672; DB 19; Length 1279;
Best Local Similarity 90.5%; Pred. No. 2.8e-163;
Matches 333; Conservative 16; Mismatches 19; Indels 0; Gaps 0;
Qy 1 TSALDTESEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVKGNHDELMK 60
Db
557 TSALDTESEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVKGNHDELMK 616
Qy 61 EKGIFYKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRSTRRSIHAPQG 120
Db 617 EKGIFYKLVMTQTAGNEVELENADESKSEIDALEMSNDSRSLIKRSTRRSVRGSQA 676
Qy 121 QDRKLGTKEDLNENVPVPSFWRILKLNSTWEPYFVGIFCAIINGLQPAFISFIISRIIG 180
Db 677 QDRKLSTKEALDESIPPVSFWRIMKLNLTWEPYFVGIFCAIINGLQPAFISFIISRIIG 736
Qy 181 IFTREDDEPETKROISNMFSVLFLVLGIISFTFFLQGTFTGKAGEILTTLKRLRYMVFERSML 240
Db 737 VFTRIDDEPETKRONSLFSLFLALGIISFTFFLQGTFTGKAGEILTTLKRLRYMVFERSML 796
Qy 241 RQDVSWFDDKNTTGALTRLANDAAQVKGAGISRLAVITONIANLGTGIIISLIYQWL 300
Db 797 RQDVSWFDDKNTTGALTRLANDAAQVKGAGISRLAVITONIANLGTGIIISLIYQWL 856
Qy 301 TLLLLAIYPIITAIAGVVEKMLSGOALKDKKELEGAGKIAEATENFTVVSLSLTOEQKFE 360
Db 857 TLLLLAIYPIITAIAGVVEKMLSGOALKDKKELEGAGKIAEATENFTVVSLSLTOEQKFE 916
Qy 361 HMYAQS LQ 368
Db 917 HMYAQS LQ 924

RESULT 12
AAB81067
ID AAB81067 standard; Protein; 1279 AA.
XX
AC AAB81067;
XX
DT 25-JUN-2001 (first entry)
XX
DE Human P-glycoprotein SEQ ID 6.
XX
KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRI;
KW efflux pump; human.
XX
OS Homo sapiens.
XX
PN WO200123565-A1.
XX
XX 05-APR-2001.
PD
XX 28-SEP-2000; 2000WO-US26592.
PF
XX 28-SEP-1999; 99US-0156921.
PR
XX 12-OCT-1999; 99US-0158818.
PR
XX (GENT-) GENTEST CORP.
PA
XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
PI WPI; 2001-316136/33.
XX
DR Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
XX (PGP) and homologous PGP polypeptides are useful for predicting
CC bioavailability of compound and increasing PGP transporter activity in
CC cell -
XX
XX Claim 6; Page 71-73; 84pp; English.
PS
XX This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDRI is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents a human
CC P-glycoprotein.
XX
XX Sequence 1279 AA;
SQ
Query Match 91.1%; Score 1672; DB 22; Length 1279;
Best Local Similarity 90.5%; Pred. No. 2.8e-163;
Matches 333; Conservative 16; Mismatches 19; Indels 0; Gaps 0;
Qy 1 TSALDTESEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVKGNHDELMK 60
Db 557 TSALDTESEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVKGNHDELMK 616
Qy 61 EKGIFYKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRSTRRSIHAPQG 120
Db 617 EKGIFYKLVMTQTAGNEVELENADESKSEIDALEMSNDSRSLIKRSTRRSVRGSQA 676
Qy 121 QDRKLGTKEDLNENVPVPSFWRILKLNSTWEPYFVGIFCAIINGLQPAFISFIISRIIG 180
Db 677 QDRKLSTKEALDESIPPVSFWRIMKLNLTWEPYFVGIFCAIINGLQPAFISFIISRIIG 736
Qy 181 IFTREDDEPETKROISNMFSVLFLVLGIISFTFFLQGTFTGKAGEILTTLKRLRYMVFERSML 240
Db 737 VFTRIDDEPETKRONSLFSLFLALGIISFTFFLQGTFTGKAGEILTTLKRLRYMVFERSML 796

QY 241 RDVSWFDDLNKNTTGALTRLANDAAQVKGAGSRSLAVITONIANLGTGIIISLIYQWL 300
 DB 797 RDVSWFDDPNKNTTGALTRLANDAAQVKGAGSRSLAVITONIANLGTGIIISLIYQWL 856
 QY 301 TLLLLAIVPIIAAGVVMKMLSGQALKDKKKELEGAGKIATEAIENFRTVVSLLTQEQKFE 360
 DB 857 TLLLLAIVPIIAAGVVMKMLSGQALKDKKKELEGAGKIATEAIENFRTVVSLLTQEQKFE 916
 QY 361 HMYAQSILQ 368
 DB 917 HMYAQSILQ 924
 RESULT 13
 AAEE00307
 ID AAE00307 standard; Protein; 1279 AA.
 XX
 AC AAE00307;
 XX
 DT 13-JUN-2001 (first entry)
 XX
 DE Human P-glycoprotein (PGP) #2.
 XX
 KW Human; P-glycoprotein; PGP; multidrug transporter; MDRI;
 KW drug bioavailability; transgenic animal; genetic model.
 XX
 OS Homo sapiens.
 XX
 PN W0200123540-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26767.
 XX
 PR 28-SEP-1999; 99US-0156510.
 XX
 PA (GENT-) GENTEST CORP.
 XX
 PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 XX WPI; 2001-235373/24.
 XX
 PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors -
 XX
 PS Claim 16; Page 80-83; 111pp; English.
 XX
 CC The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDRI) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterised by PGP activity, by
 CC reducing or increasing PGP activity in a cell. PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is human P-glycoprotein (PGP). The
 CC PGP enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.
 XX
 SQ Sequence 1279 AA;
 Query Match 91.1%; Score 1672; DB 22; Length 1279;
 Best Local Similarity 90.5%; Pred No. 2.8e-163;
 Matches 333; Conservative 16; Mismatches 19; Indels 0; Gaps 0;
 QY 1 TSALDTESEAVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 60

DB 557 TSALDTESEAVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 616
 QY 61 EKGIFYKLVMTQTRGNEIDLENATGESKSDALENPKDSGLIKRSTRSIIHAPOG 120
 DB 617 EKGIFYKLVMTQTRGNEIDLENATGESKSDALENPKDSGLIKRSTRSIIHAPOG 676
 QY 121 ODRKLGTKEDLNENPVPSFWRILKLNSTWPFYVVGIFCAIINGLQPAFISIFSRIG 180
 DB 677 ODRKLGTKEDLNENPVPSFWRILKLNSTWPFYVVGIFCAIINGLQPAFISIFSRIG 736
 QY 181 IFTREDDEPETKROISNMESVLFVLGLISFITEFLOGTFGKAGEILLTKRLRYWVFRSML 240
 DB 737 VFTRIDDEPETKRONSLFSLFLALGLISFITEFLOGTFGKAGEILLTKRLRYWVFRSML 796
 QY 241 RDVSWFDDLNKNTTGALTRLANDAAQVKGAGSRSLAVITONIANLGTGIIISLIYQWL 300
 DB 797 RDVSWFDDPNKNTTGALTRLANDAAQVKGAGSRSLAVITONIANLGTGIIISLIYQWL 856
 QY 301 TLLLLAIVPIIAAGVVMKMLSGQALKDKKKELEGAGKIATEAIENFRTVVSLLTQEQKFE 360
 DB 857 TLLLLAIVPIIAAGVVMKMLSGQALKDKKKELEGAGKIATEAIENFRTVVSLLTQEQKFE 916
 QY 361 HMYAQSILQ 368
 DB 917 HMYAQSILQ 924
 RESULT 14
 AAP70452
 ID AAP70452 standard; Protein; 1280 AA.
 XX
 AC AAP70452;
 XX
 DT 21-MAY-1991 (first entry)
 XX
 DE Sequence encoded by human multi-drug resistance-1 (mdr1) cDNA
 DE from clones lambda-HDR10,5 and 104.
 XX
 KW Chemo-therapy resistant tumour cell; P-glycoprotein.
 XX
 OS Homo sapiens.
 PN W08705943-A.
 XX
 PD 08-OCT-1987.
 XX
 PF 26-MAR-1987; 87WO-US00758.
 PR 01-AUG-1986; 86US-0892575.
 PR 28-MAR-1986; 86US-0845610.
 XX
 PA (UNII) UNIV OF ILLINOIS.
 XX
 PI Roninson IB, Pastan IH, Gottesman MM;
 XX WPI; 1987-291656/41.
 DR N-P5DB; AAN70752.
 XX
 PT DNA for multi-drug resistance in human cells - used to detect
 PT chemotherapy-resistant tumour cells and for producing
 PT polypeptide(s) for diagnosis and therapy
 XX
 PS Claim 4(a); Table 5, pp30-39; 61pp; English.
 XX
 CC The human multi-drug resistant KB carcinoma cell lines were used as
 CC the source of the mdr1 gene nucleic acid sequences (AAN70751). To
 CC obtain cDNA clones of the mdr1 gene (AAN70752), poly (A) and RNA was
 CC used. Analysis of the AA sequence presented in (AAN70752) indicates
 CC that the mdr1 gene product is likely to be a transmembrane protein.
 CC The presence of transmembrane domains and potential glycosylation
 CC sites is consistent with the mdr1 protein being related to the
 CC P-glycoprotein.

```
XX SQ Sequence 1280 AA;
Query Match 91.1%; Score 1672; DB 8; Length 1280;
Best Local Similarity 90.5%; Pred. No. 2.8e-163;
Matches 333; Conservative 16; Mismatches 19; Indels 0; Gaps 0;

QY 1 TSALDTESEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 60
DB 558 TSALDTESEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 617
QY 61 EKGIIYFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRRSTRSIIHAPQG 120
DB 618 EKGIIYFKLVMTQTAGNEVELENADESKSEIDALEMSNDSSRLIKRRSTRSIVRGSOA 677
QY 121 QDRKLTGKEDLNENVPVSWRILKLNSTWEPYFVVGIFCAIINGGLQPAFIIIFSRIIG 180
DB 678 QDRKLTGKEDLNENVPVSWRILKLNSTWEPYFVVGIFCAIINGGLQPAFIIIFSRIIG 737
QY 181 IFTREDEPETKROISNMFSVLFLVGLIISFTFFLQGTGKAGEILTCKRLRYMVFRSML 240
DB 738 VFTRIDDPETKRONSLFSLFLALGIISFTFFLQGTGKAGEILTCKRLRYMVFRSML 797
QY 241 RODVSFDDLNKNTGALTTRLANDAAQVKGAGISRLAVITONIANLGTGIIISLIYGMOL 300
DB 798 RODVSFDDPNKNTGALTTRLANDAAQVKGAGISRLAVITONIANLGTGIIISLIYGMOL 857
QY 301 TLLLLAIVIPPIAIAAGVEMKMLSGQALKDKKEGAGKIATEAENFRTVVSILTQEQKFE 360
DB 858 TLLLLAIVIPPIAIAAGVEMKMLSGQALKDKKEGAGKIATEAENFRTVVSILTQEQKFE 917
QY 361 HMYAQSLQ 368
DB 918 HMYAQSLQ 925

RESULT 15
AAR63624
ID AAR63624 standard; Protein; 1280 AA.
XX AC AAR63624;
XX DT 21-JUN-1995 (first entry)
XX DE Human p-glycoprotein encoded by the multidrug resistance gene MRD-1.
XX KW Human; tyrosinase; compound; hair; follicle; liposome; lipophilic;
XX KW lipophilic; stratum corneum; melanin; hair dye; glycoprotein; alopecia;
XX KW chemotherapy; transforming growth factor; growth stimulant; aromatase;
XX KW cyclosporin A; macromolecule; polymer.
XX OS Homo sapiens.
XX PN W09422468-A.
XX PD 13-OCT-1994.
XX PF 01-APR-1994; 94WO-US03634.
XX PR 02-APR-1993; 93US-0041553.
XX PR 13-JAN-1994; 94US-0181471.
XX PA (ANTI-) ANTICANCER INC.
XX PI Li L, Lishko VK;
XX WI: 1994-332816/41.
XX N-PSDB; AAQ72872.
XX Lipsomes for delivering protein, nucleic acid etc. to hair
XX follicles - e.g. to restore hair colour, prevent hair loss during
XX chemotherapy, stimulate hair growth etc.
```

```
PS XX The amino acid sequence of a human p-glycoprotein encoded by the
CC multidrug resistance gene MRD-1. This is an example of a compound which
CC can be delivered to hair follicles via a novel liposome composition.
CC The liposomes are comprised of a lipophilic or lipophobic compound which
CC will selectively target the hair follicle (via the stratum corneum)
CC without damaging or unwanted effects on cells outside the follicle.
CC Compounds e.g. p-glycoprotein, can be delivered to the hair follicle for
CC treatment of chemotherapy-induced alopecia. Other compounds targeted at
CC hair follicles can include: tyrosinase (or the DNA encoding it -
CC AAQ72871), melanin or hair dyes (to restore hair colour or condition);
CC human transforming growth factor-alpha (AAQ72873) (for reversal of wavy
CC hair); hair growth stimulants (e.g. cyclosporin A or aromatase) or
CC antisense sequences. The method allows compounds (e.g. macromolecules
CC or polymers), which would not normally reach the hair follicles, to be
CC delivered to these target areas.
XX SQ Sequence 1280 AA;
Query Match 91.1%; Score 1672; DB 15; Length 1280;
Best Local Similarity 90.5%; Pred. No. 2.8e-163;
Matches 333; Conservative 16; Mismatches 19; Indels 0; Gaps 0;

QY 1 TSALDTESEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 60
DB 558 TSALDTESEAVVQVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVEKGNHDELMK 617
QY 61 EKGIIYFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRRSTRSIIHAPQG 120
DB 618 EKGIIYFKLVMTQTAGNEVELENADESKSEIDALEMSNDSSRLIKRRSTRSIVRGSOA 677
QY 121 QDRKLTGKEDLNENVPVSWRILKLNSTWEPYFVVGIFCAIINGGLQPAFIIIFSRIIG 180
DB 678 QDRKLTGKEDLNENVPVSWRILKLNSTWEPYFVVGIFCAIINGGLQPAFIIIFSRIIG 737
QY 181 IFTREDEPETKROISNMFSVLFLVGLIISFTFFLQGTGKAGEILTCKRLRYMVFRSML 240
DB 738 VFTRIDDPETKRONSLFSLFLALGIISFTFFLQGTGKAGEILTCKRLRYMVFRSML 797
QY 241 RODVSFDDLNKNTGALTTRLANDAAQVKGAGISRLAVITONIANLGTGIIISLIYGMOL 300
DB 798 RODVSFDDPNKNTGALTTRLANDAAQVKGAGISRLAVITONIANLGTGIIISLIYGMOL 857
QY 301 TLLLLAIVIPPIAIAAGVEMKMLSGQALKDKKEGAGKIATEAENFRTVVSILTQEQKFE 360
DB 858 TLLLLAIVIPPIAIAAGVEMKMLSGQALKDKKEGAGKIATEAENFRTVVSILTQEQKFE 917
QY 361 HMYAQSLQ 368
DB 918 HMYAQSLQ 925

Search completed: November 6, 2002, 18:39:20
Job time : 14.0475 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:45:15 ; Search time 1092.74 Seconds
(without alignments)
7047.385 Million cell updates/sec

Title: US-09-672-725C-6
Perfect score: 1836
Sequence: 1 TSALDTSESEAVGVQALDKAR.....TVVSLTQEQKFEHMYAQSILQ 368

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_SPOOL/US09672725/runat_04112002_124341_17977/app_query.fasta_1.6435
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725@cgn_1_11409@runat_04112002_124341_17977 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb.ba: *
2: gb.htg: *
3: gb.in: *
4: gb.om: *
5: gb.ov: *
6: gb.pat: *
7: gb.ph: *
8: gb.pl: *
9: gb.pr: *
10: gb.ro: *
11: gb.sts: *
12: gb.sy: *
13: gb.un: *
14: gb.vi: *
15: em.ba: *
16: em.fun: *
17: em.hum: *
18: em.in: *
19: em.mu: *
20: em.on: *
21: em.or: *
22: em.ov: *
23: em.pat: *
24: em.ph: *
25: em.pl: *
26: em.ro: *
27: em.sts: *
28: em.un: *

29: em.vi: *
30: em.htg_hum: *
31: em.htg_inv: *
32: em.htg_other: *
33: em.htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1836	100.0	1107	4	AF092810 Canis fam
2	1836	100.0	1107	6	AX105061 Sequence
3	1808	98.5	3934	4	CFA419568 Canis fam
4	1808	98.5	4045	12	AF269224 Synthetic
5	1808	98.5	4279	6	AX105057 Sequence
6	1808	98.5	4279	6	AX105078 Sequence
7	1808	98.5	4279	6	AX105080 Sequence
8	1808	98.5	4279	6	AX105082 Sequence
9	1800	98.0	4317	4	AF045016 Canis fam
10	1800	98.0	4317	6	AX105059 Sequence
11	1764	96.1	2005	4	AB066299 Canis fam
12	1700	92.6	3489	4	AB029153 Canis cat
13	1678	91.4	4186	6	AX108654 Sequence
14	1678	91.4	4195	6	AX108656 Sequence
15	1672	91.1	2726	6	I33621 Sequence 1
16	1672	91.1	3860	6	AX322787 Sequence
17	1672	91.1	3860	6	AX322789 Sequence
18	1672	91.1	3988	6	AX024454 Sequence
19	1672	91.1	4192	9	AF016535 Homo sapi
20	1672	91.1	4264	6	AR051647 Sequence
21	1672	91.1	4264	6	AR051650 Sequence
22	1672	91.1	4646	6	AX336420 Sequence
23	1672	91.1	4646	6	AX336708 Sequence
24	1672	91.1	4646	6	I49610 Sequence
25	1672	91.1	4646	6	HUMDR1
26	1672	91.1	4669	6	I08557 Sequence 3
27	1672	91.1	6505	6	AR028671 Sequence
28	1672	91.1	8630	6	AX012320 Sequence
29	1672	91.1	8630	6	AX012321 Sequence
30	1672	91.1	9318	6	AR028672 Sequence
31	1669	90.9	4378	6	E02326 Multidrug r
32	1653	90.0	4669	6	AR055785 Sequence
33	1644	89.5	4669	6	AR091275 Sequence
34	1589	86.5	3858	4	OAU78609 Ovis aries
35	1589	86.5	3987	10	CRUPGP1185 Chinese ham
36	1589	86.5	4296	10	CRUPGP1 Chinese ham
37	1589	86.5	4304	10	CRUPGP1165 Chinese ham
38	1584	86.3	3682	10	MUSPGLX1A Mouse phosph
39	1584	86.3	4356	10	MUSMDRAA Mouse multi
40	1584	86.3	4788	6	AX322793 Sequence
41	1584	86.3	4924	10	MUSMDRIA Mouse P-gly
42	1575	85.8	4927	10	AF257746 Rattus no
43	1566	85.3	4323	10	AF286167 Rattus no
44	1426.5	77.7	4233	6	AR123273 Sequence
45	1418.5	77.3	4254	10	RATMDRM Rat mdr mrn

ALIGNMENTS

RESULT 1 AF092810
LOCUS AF092810
DEFINITION Canis familiaris multidrug resistance protein MDRL mRNA, partial
ACCESSION AF092810
VERSION AF092810.1
KEYWORDS cds.
SOURCE dog.
1107 bp mrna linear MAM 04-OCT-1998

ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 1107)
AUTHORS Steingold,S.F., Sharp,N.J., McGahan,M.C., Hughes,C.S., Dunn,S.E.
and Page,R.L.
TITLE Characterization of canine MDRI mRNA: its abundance in drug
resistant cell lines and in vivo
JOURNAL Anticancer Res. 18 (1A), 393-400 (1998)
MEDLINE 98229528
REFERENCE 2 (bases 1 to 1107)
AUTHORS Steingold,S.F., Sharp,N.J., McGahan,M.C., Hughes,C.S., Dunn,S.E.
and Page,R.L.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1998) College of Veterinary Medicine, North
Carolina State University, 4700 Hillsborough St, Raleigh, NC 27606,
USA
FEATURES
source Location/Qualifiers
1..1107
/organism="Canis familiaris"
/db_xref="taxon:9615"
CDS <1..>1107
/note="similar to Homo sapiens MDRI"
/codon_start=1
/product="multidrug resistance protein MDRI"
/protein_id="AAC62505.1"
/db_xref="GI:3694968"
/translation="PSALDTESEAVQVALDKARKGRTTIVIAHRLSVRNADVIAGF
DGVIVKRNHDELKWKGYFKLVTMQRGNEIDLENATGESKESDALEMSPKDSG
SSLIKRRSHIAPQGDRLKEDLNENPVPSFWRLIKLMSDEWPTFVVGIFC
AINGLOPAFIIFRSIIGTRFDEDPETKQISNMFSLVFLGIISFTIFFLQGF
TFGKAGEILTKFRVYFMSRLRQDVSWFDLTKTGALTTLANDAAQVKAIGSRL
AVTQNIANLGIIISLYIYWGQLTLLIIVPIAIAGVVMKLSQALKDKKKELE
GAGKATAEINFRVSLTQKFEHMYAQSLOV"
BASE COUNT 341 a 200 c 256 g 310 t
ORIGIN
Alignment Scores:
Pred. No.: 8 6e-158 Length: 1107
Score: 1836.00 Matches: 368
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-09-672-725C-6 (1-368) x AF092810 (1-1107)
QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArg 20
Db 1 ACGTCAGCTGGACACTGAAGTGAAGCAGTGGTTCAGGTGGCCCTGGGATAAGGCCAGA 60
QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 61 AAAGCCGGACTACCATGTGATAGCTCATCGTTGTCTACAGTTCGTAATGCCGATGTC 120
QY 41 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisaspGluLeuMetLys 60
Db 121 ATTGCTGTTTTGTATGATGGAGTCATTGTGGAGAAAGAAATCATGAACTCATGAAA 180
QY 61 GluLysGlyIleTyrrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
Db 181 GAGAAAGGCATTATCTCAAACTTGTCCAAATGCAGACAGAGGAATGAAATGACTTGA 240
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 241 GAAATGCCACTGGTGAATCAAAAGTGAAGTCAATGCCAGACAGAGGAATGAAATGACTTGA 300
QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
Db 301 TCAGGGTCCAGTTTAATAAAAAAGAGATCAACTCCGAGGAGTATACATGCACCAAGGC 360
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 140

Db 361 CAAGACACAAAGCTTGTTGTAACAAAGAGAGACTTGTAATGAGAAATGTACCTCCAGTTTCCTTC 420
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrrPheValValGlyIlePheCys 160
Db 421 TGGAGGATTCGAAGCTCAACTCAACTGAATGGCTTATTTTGTGGTTGGTATATTTGT 480
QY 161 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 481 GCTATTATAAACGGAGGCTGCAGCCAGCAATTTTCAATAATATTTTCAAGGATTATAGG 540
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 541 ATCTTTTACCCGAGATGAGATCCTGAAACAAACAGACAGATTAGTAAACATGTTCTGTA 600
QY 201 LeuPheLeuValLeuGlyIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 601 TTGTTCTAGTCCCTGGAATATTTCTTTTATTACATTTTCTTCAGGGTTTCCACATTT 660
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrrMetValPheArgSerMetLeu 240
Db 661 GGCAAGCTGGAGAGATCCTCACTAAGCGGCTCGATACATGGTTTTCAGATCCATGCTG 720
QY 241 ArgGlnAspValSerTrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 721 AGACAGGATGTCAGCTGGTTTGTATGACCTTAAACACCACTGGAGCATTGACCCAGG 780
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 781 CTTGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCTGCTATTACC 840
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrrGlyTrpGlnLeu 300
Db 841 CAGAATATAGCAAAATCTTGGGACAGGCATTATATATCTTAACTATGTTGGCAATTA 900
QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
Db 901 ACATTTTACTCTTAGCAATTGTACCCATCATTCGAATAGCAGGAGTTGTTGAAATGAAA 960
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 961 ATGTTGCTCTGGACAAGCAGCTGAAAGATAAGAAAGAGCTAGAGGAGCTGGGAAGATTGCT 1020
QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGlnLysPheGlu 360
Db 1021 ACAGAAGCCATCGAAACTTCGAACTGTGTTCTTTGACTCAGGACGAGAGTTTGAA 1080
QY 361 HisMetTyrrAlaGlnSerLeuGln 368
Db 1081 CACATGTATGCACAGAGTTTGCAG 1104
RESULT 2
AX105061
LOCUS 1107 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 5 from Patent WO0123540.
ACCESSION AX105061
VERSION AX105061.1 GI:13921213
KEYWORDS
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 1107)
AUTHORS Stocker,P.J., Steimel-Crespi,D.T., Crespi,C.L., Reif,T.C. and
Patten,C.J.
TITLE P-glycoproteins and uses thereof
JOURNAL Patent: WO 0123540-A 5 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
source Location/Qualifiers
1..1107
/organism="Canis familiaris"
/db_xref="taxon:9615"
CDS <1..>1107
/note="unnamed protein product"

Best Local Similarity: 98.64%				Mismatches: 2	
Query Match: 98.47%				Indels: 0	
DB: 4				Gaps: 0	
US-09-672-725C-6 (1-368) x CFA419568 (1-3934)					
Qy	1	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	20		
Db	1680	ACGTCAGCTCTGGACACATGAAGTGAACAGCTGGTTTCAGGTGGCCCTGGATAGGCCAGA	1739		
Qy	21	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	40		
Db	1740	AAAGCGCGCATACCATTTGTGATAGCTCATCGTTTGTCTACAGTTCTGTAATGCCGATGTC	1799		
Qy	41	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	60		
Db	1800	ATTGCTGGTTTTGATGATGGAGTCATTTGGAGAAAGGAATCATGATGAACATCATGAAA	1859		
Qy	61	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu	80		
Db	1860	GAGAAGGGCATTTACTTCAAACTTGTCACAAATGCAGACAAGAGAAATGAAATGTAGTTA	1919		
Qy	81	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	100		
Db	1920	GAATAATGCCACTGGTGAATCCAAAAGTGAAGTGTGCTTGGAAATGTCCTCAAAAGAT	1979		
Qy	101	SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly	120		
Db	1980	TCAGGGTCCAGTTTATAAAGAGAGATCAACTCGCAGGAGTATACATGCCACCAAGGC	2039		
Qy	121	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	140		
Db	2040	CAAGACAGAAGCTTGGTACAAAGAGGACTTGAATGAGATGTACTCCAGCTTTCCTTC	2099		
Qy	141	TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	160		
Db	2100	TGGAGGATTCGAAAGCTGAAGCTCAACTGAAAGGCCCTTATTTGTGGTTGGTATATTTGT	2159		
Qy	161	AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	180		
Db	2160	GCTATTATAACGGAGCGCTGCAACAGCATTTTCAATAAATATTTTCAAGGATTATAGG	2219		
Qy	181	IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal	200		
Db	2220	ATCTTTACCCGAGATGAGATCTGAAACAAACACGACAGAAATAGTAACATGTTTCTGTA	2279		
Qy	201	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	220		
Db	2280	TTGTTTCTAGTCTGGAATATTTCTTTTATTTACATTTTTTCTCCAGGGCTTCACATTT	2339		
Qy	221	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	240		
Db	2340	GGCAAAAGCTGGGAGATCCTCATAAGCGGCTTCGATACATGTTTTCAGATCCATGCTG	2399		
Qy	241	ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg	260		
Db	2400	AGACAGGATGTCAGCTGGTTTGTATGACCCCTAAAACACCACTGGAGCATTTGACAAACAGG	2459		
Qy	261	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	280		
Db	2460	CTTGCCAATGATGGGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTGTCATTACC	2519		
Qy	281	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	300		
Db	2520	CAGAAATATAGCAATCTTGGACAGGCAATATATATCTTAAATCTATGTTGGCAATTA	2579		
Qy	301	ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	320		
Db	2580	ACACTTTTACTCTTAGCAATTTGTACCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAAA	2639		
Qy	321	MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla	340		
Db	2640	ATGTTGCTGGACAAAGCACTGAAAGATAGAAAGAGCTAGAAAGAGCTGGGAGAGTTGCT	2699		

Qy	341	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGlnLysPheGlu	360	
Db	2700	ACAGAAGCCATCGAAACTTCGGAACGTGTTCTTTGACTCGGGAGCAGAAGTTTGA	2759	
Qy	361	HisMetTyrAlaGlnSerLeuGln	368	
Db	2760	TACATGTATGCACAGAGTTTGC	2783	
RESULT 4				
AF269224		4045 bp	DNA	linear
LOCUS	SYN 04-JUL-2000			
DEFINITION	Synthetic construct Canis familiaris his-tagged-multidrug resistance glycoprotein gene, complete cds.			
ACCESSION	AF269224			
VERSION	AF269224.1	GI:8926216		
KEYWORDS	synthetic construct.			
SOURCE	synthetic construct			
ORGANISM	artificial sequence.			
REFERENCE	1 (bases 1 to 4045)			
AUTHORS	Roulet,A.C., Puel,O., Gesta,S., Alvinerie,M. and Pineau,T.J.			
TITLE	Analysis of dog MDR1 p-glycoprotein			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 4045)			
AUTHORS	Roulet,A.C., Puel,O., Gesta,S., Alvinerie,M. and Pineau,T.J.			
TITLE	Direct Submission			
JOURNAL	Submitted (19-MAY-2000) Pharmacology, INRA, 180 Chemin de Tournepieuille, Bp3, Toulouse 31391, France			
FEATURES	Location/Qualifiers			
source	1..4045	/organism="synthetic construct"		
		/db_xref="taxon:32630"		
		/note="original mRNA isolated from Beagle dog colon"		
	101..3964	/codon_start=1		
		/transl_table=11		
		/product="his-tagged-multidrug resistance glycoprotein MDR1"		
		/protein_id="AAF81747.1"		
		/db_xref="GI:8926217"		
		/translation="MHHHHDDPEGGRKSAEKNFWKMGKSKKNEKKKKPTVSTFA		
		MFYSNWLDRYLMVGTMAAIHGAALPLMLVFGNMTDSFANAGISRNKTPFVINE		
		SINTNTHQFINHLEEMTYAYVYSGIAGVLYAAVYQVSWFCLAAAGROLTKRKOFF		
		HAINRQEIQFVDHVGELNTRLRDDVSKINEGIDKIGMFQSIATFTFTGLVGTFR		
		GKMLTVLAISPVLGSAAIWAKILSSFTDKELLATYAKAGAAVEVLAIRVIAFG		
		GQKLELYRNKLEAKRIGIKKATANI SIGAFLLIYALAYALAWYGVLSLSEY		
		SIGOVLTVFTSVLIGAFSGQASPIEAFANARGAAVEIFKIIDNKPSIDSYSGSHK		
		PDNTGNLEPKNVHFSYPSRKEVKILKLNKVQSGOTVALVNSGCGKSTVOLMQR		
		LYDPTGMVCDGODIPTINVRHLREITGVVSOEPLVATTIAENTRYGRENVTMBEI		
		EKAYKANADYFIMKLPNKFTLGERGAOLSGQOKRIARALVRNPKILLDEAT		
		SALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVTAGDDGVIVKGNHDELM		
		KEKGIYFLVTMQTRNEIELENATGESKSEDALEMSPKDSGLIKRKRSTRRSHIA		
		POGDRKLGTKEDLNENVPVPSFWRLKLNSTPEYVFGVIFCAINGLOPAPSIIF		
		SRIIGIFTRDSDPETKQNSNMFVLVLGIIISFTIFFLQGTFTKRGAILTKRLRY		
		MYFMSLRQDYSWFDDPKNTGALTTLANDAAQVKAIGSLAVITONTANLGTII		
		ISLYIGQTLTLLLAIVPITAIGVVEKMLSGQALKKELEGAGKIATEIENFT		
		VYSLTRQKFEYMTAQSLQVPYRNSLRKHAHFGVFSITQAMMYFVAGGCFRGAVLY		
		ANEFNFDVLVFSALVFGAMAVQVSSFADYAKAVSAAHVIMIEIKSPSIDTS		
		PHGLKPNLTLEGNVTNEVFNYPTRDPDIPVLQGLSLVKKQTALVYSGGCKSTV		
		OLLRRFYDPLAGSVLIDGKEIKHLNVQWLAHLGIVSQEPTLFDCTSAENTAYGNSR		
		VVSHEETMQAAKEANIHFIFETLPEKYNTRVKGDTOLSGGOKORIAARALVRQPH		
		LLDDEATSAUTESKVVQEALEDKARERTCIVIAHRLSTIQNALIVVFNQNGKVEH		
		GTHOOLLAQKGIYFMSISVQAGAAK"		
BASE COUNT	1171 a	850 c	986 g	1038 t
ORIGIN				
Alignment Scores:	1.55e-154	Length:	4045	
Pred. No.:	1808.00	Matches:	363	
Score:	99.46%	Conservative:	3	
Percent Similarity:	98.64%	Mismatches:	2	
Best Local Similarity:	98.47%	Indels:	0	
Query Match:				

DB:	12	Gaps:	0
US-09-672-725C-6 (1-368) x AF269224 (1-4045)			
QY	1	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	20
DB	1796	ACGTGACGCTGGACACTGAAAGTGAACAGTGGTTTCAGGTGGCCCTGGATAAGGCCAGA	1855
QY	21	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	40
DB	1856	RAAGCCCGACTACCATTTGTGATGACTCATCGTTTGTACAGTTCGTAATGCCGATGTC	1915
QY	41	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	60
DB	1916	ATTGCTGTTTGTGATGGAGTCATTGTCGAGAAAGGAATCATGATGAACCTCATGAA	1975
QY	61	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu	80
DB	1976	GAGAAAGGCACTTACTTCAAACTTGTCAATGCAGACAGAGGAAATGAAATTCAGTTA	2035
QY	81	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	100
DB	2036	GAAATGCCACTGGTGAATCCAAAGTGAAAGTATGCTTGGAAATGCTCCAAAAGAT	2095
QY	101	SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly	120
DB	2096	TCAGGGTCCAGTTTAAATAAAAGAAAGATCAACTCGCAGGAGTATACATGCACCACAGG	2155
QY	121	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	140
DB	2156	CAAGACAGAAAGCTTGTGTACAAAAGAGGACTTGAATGAGAAATGATACCTCCAGTTT	2215
QY	141	TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	160
DB	2216	TGGAGGATTCGAGCTGAACTCAACTGAATAGGCCCTATTTTTGGTGGTATATTGT	2275
QY	161	AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	180
DB	2276	GCTATTATAACGGAGGCTGCAACACGAGCTTTTCAATAATATTTCACGATATATAGG	2335
QY	181	IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal	200
DB	2336	ATCTTTACCCGATGAGGATCCTGAAACAAACGACAGAAATAGTAACATGTTTCTGTA	2395
QY	201	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	220
DB	2396	TTGTTTCTAGTCCTGGAATATTCTTTTATTACATTTTCTCCAGGGCTTCACATTT	2455
QY	221	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	240
DB	2456	GGCAAGCTGGGAGATCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTG	2515
QY	241	ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg	260
DB	2516	AGACAGATGTCAGCTGGTTGATGACCCCTAAACACCACTGGAGCATTCGACACGAG	2575
QY	261	LeuAlaAsnAspAlaLagInValLysGlyValaIleGlySerArgLeuAlaValIleThr	280
DB	2576	CTTGCCAATGATCGGCTCAAGTTAAAGGGCTATAGTTCAGGCTTCTGTCATTACC	2635
QY	281	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeu	300
DB	2636	CAGAATATAGCAATCTTGGACAGCACTATTATATATATATATATATATATATATATAT	2695
QY	301	ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	320
DB	2696	ACACTTTTACTCTAGCAATGTACCAATGTACCAATGTACCAATGTACCAATGTACCAAT	2755
QY	321	MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla	340
DB	2756	ATGTTGCTGGCAAGCACTGAAAGATAGAAGAGCTAGAGGAGCTGGGAAGATGCT	2815
QY	341	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluLysPheGlu	360

DB	2816	ACAGAACCCATCGAAACTCCGAACTGTTGTTCTTTGACTCGGAGACAGAAAGTTGAA	2875
QY	361	HisMetTyrAlaGlnSerLeuGln	368
DB	2876	TACATGTATCACAGAGATTTGCAA	2899
RESULT 5.			
AX105057	AX105057	4279 bp	DNA linear PAT 30-APR-2001
LOCUS	Sequence 1 from Patent WO0123540.		
DEFINITION	AX105057		
ACCESSION	AX105057.1		
VERSION	GI:13921209		
KEYWORDS	dog.		
SOURCE	Canis familiaris		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
AUTHORS	Stocker, P.J., Steimel-Crespi, D.T., Crespi, C.L., Reif, T.C. and Patten, C.J.		
TITLE	P-glycoproteins and uses thereof		
JOURNAL	Patent: WO 0123540-A 1 05-APR-2001;		
FEATURES	GENTEST CORPORATION (US)		
Source	Location/Qualifiers		
	1..4279		
	/organism="Canis familiaris"		
	/db_xref="taxon:9615"		
	17..3862		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC37718.1"		
	/db_xref="GI:13921210"		
CDS	/translation="MDPEGGRKGSAEKNFKMKKSKKNEKKKPKTVSTFAMFRYSN		
	WLDRLYLVTGMAAIIHGAAALFMMLVFNGNMTDSFANAGISRNKTFPVIINESITNNT		
	OHFINHLEEMTTVAYYSGIGAGVLVAAYIQVFSGLAAGROILKIRKOPFHAIIMRQ		
	EIGWFDVHDVGLNTRLDYDSKNEGIDKIGMEFHSIATFTFGFVGRGKLLI		
	VILATSPVLGLSAIWAIISSFTDKELLAYAKAGAVAEVLAIRVIAFGGOKKEL		
	ERYNNLEEKAGIGIKKAITANISGAFLIILYASIALAFWIGLSLSSSEYSIGQVL		
	TVFVSLIGAFSIGQSPSIEAFANARGAAEYEIFKIDNKPSIDSYSKSGHKPDNIKG		
	NLEFKNVHFSYPSRKEIKLGLNLKVSGQTVALVNSGCGKSTVQLMQLDYDPTD		
	GNVICDQDITINVRLHREITGVVSOBPLVFTTIAENIRYGRNVTMDIEKAVKE		
	ANAYDFIMKLPKFTLVGERGAOLSGGOKORIAARALVNRPKILLDEATSLDTE		
	SEAVVOVALDKARKGRTTIVIAHRLSTVRNADVAGDDGVIVEKGNHDELMKEGIIY		
	PKLYTMQTRGNEIELENATGESKSESDALEMSPKDSSGLKRRSTRSRSHAPOQQR		
	KTRDEDPENPVPSVRILKLNSTEMPYEVGVFCALINGGLQAPAFIIFSRILGI		
	FLTRDEDPKTRQNSMFSVLEVLGIISFTIFLQGTGKAGEILTKRLRYMVRSM		
	LRQDYSWFDPPKNTTGALTRLANDAQAQKAGSLAVITQNIANLGTIIISLIYG		
	WQLTLLILAIPIIAIAGVEMKMLSGQALKDKKELEGAGKIAATEIENPRTVSLTR		
	EOKFEYMAQSLQVYRNLRAHIFGVFSITQAMMYFSVAGCFRFGAYLVANEFMN		
	FQDVLIVPSAIVFGAMAVGVQSFAPDYAKAKVAAHVIMIEKSPDISYSPHGLAP		
	NILGNVTNEVWFNYPTRDIPVLQGLSLEVKKGQTLALVSSGCCSKSTVVQLLEFP		
	IMQAAKENIHFFIETLPKYNTRVGDKGTQSGGOKORIAARALVNRQPHILLDEA		
	TSALDTESEKVVQALDKAREGTCIVIAHRLSTIQNADLTVVFQNGKVEHGTHQQL		
	LAQKGIYFSMVSVQAGAKR"		
BASE COUNT	1294 a	834 c	1008 g 1143 t
ORIGIN			
Alignment Scores:			
Pred. No.:	1.66e-154	Length:	4279
Score:	1808.00	Matches:	363
Percent Similarity:	99.46%	Conservative:	3
Best Local Similarity:	98.64%	Mismatches:	2
Query Match:	98.47%	Indels:	0
DB:	6	Gaps:	0
US-09-672-725C-6 (1-368) x AX105057 (1-4279)			
QY	1	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	20
DB	1694	ACGTGACGCTGGACACTGAAAGTGAACAGTGGTTTCAGGTGGCCCTGGATAAGGCCAGA	1753

QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1754 ARAAGCGGACCTACCATCTGATAGCTCATCTGTTGTCTACAGTTCTGTAATGCCGATGTC 1813
QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1814 ATTGCTGGTTTGTATGATGAGTGCATTTGTGGAGAAAGGAAATCATGCAACTCATGAAA 1873
QY 61 GluLysGlyLeuThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
Db 1874 GAGAAGGGATTACTTCAAACTTGTCAAAATGTCAGACAGAGGAAATGAAATTCAGTTA 1933
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 1934 GAAATGCCACTGGTGAATCCAAAGTGAAGTGTGCTTGGAAATGCTCCCAAGAT 1993
QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
Db 1994 TCAGGCTCCAGTTTAATAAAAGAGATCAACTCGCAGAGATATACATGCACCAAGGC 2053
QY 121 GlnAspArgLysGluThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
Db 2054 CAAGACAGAAGCTTGGTACAAAGAGACTTGAATGACAAATGTACCTCCAGTTCCTTC 2113
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
Db 2114 TGGAGGATCTGAAGCTGAACCTCACTGAATGCGCTTATTTGTGGTTGTTATTTTGT 2173
QY 161 AlaIleAsnGlyLysGluGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 2174 GCTATTATAACAGGAGGCTTCAACAGCATTTTCAATAATATTTTCAAGGATTTAGGG 2233
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2234 ATCTTACCCAGATGAGGATCTCTGAACAAACGACAGAAATAGTACATGTTTCTGTA 2293
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 2294 TTGTTTCTAGCTCTTGGAAATTTATTTCTTTTATTTACATTTTCTCCAGGCGTTTCACTTT 2353
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2354 GGCAAAGCTGGGAGATCCCTACCTAGCGCTTCGATACATGTTTTCAGATCCATGCTG 2413
QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 2414 AGACAGATGTCAGCTGTTTGTATGACCCCTAAACACACCACTGGAGCAATTGCAACACG 2473
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2474 CTTGCCAATGATGGGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGTGTCATTACC 2533
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 2534 CAGAAATATAGCAATCTTTGGACAGGCAATTTATATCTTATCTATCTATGTTGGCAATTA 2593
QY 301 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
Db 2594 ACACCTTTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGAGTTGTTGAAATGAAA 2653
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 2654 ATGTGTGCTGGACAGACACTGAAAGATTAAGAAGAGAGCTAGAAGAGCTGGGAAGATTGCT 2713
QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
Db 2714 ACAGAGCCATCGAAACTTCGAACTGTTGTTCTTTGACTCGGAGCAGAAAGTTTGAA 2773
QY 361 HisMetTyrAlaGlnSerLeuGln 368
Db 2774 TACATGTATGCACAGAGTTTGCAA 2797

RESULT 6

AX105078 4279 bp DNA linear PAT 30-APR-2001
LOCUS Sequence 22 from Patent WO0123540.
DEFINITION AX105078
ACCESSION AX105078
VERSION 1
KEYWORDS AX105078.1 GI:13921228
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 4279)
AUTHORS Stocker,P.J., Steimel-Crespi,D.T., Crespi,C.L., Reif,T.C. and
Patten,C.J.
TITLE P-glycoproteins and uses thereof
JOURNAL Patent: WO 0123540-A 22 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
Location/Qualifiers
1..4279
/organism="Canis familiaris"
/db_xref="taxon:9615"
17..3862
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37721.1"
/db_xref="GI:13921229"
/translation="MDPEGKGSAAKFNWKKGKSKKNEKKKKPTVSTFAMFRYSN
WLDRLYMLVGTMAAIIHGAALPLMLVFGNMTDSFANAGISRNKTPVIVINESITNTT
QHFINLHEEMTTAYIYSGIGAGLVAAIYQVFWCLAAAGQILIKRFTGFWHAIWRQ
EIGFNDVDVDELNTRLTDVSKINIGDKIGMFFQSTATFTGTFVGTGKWLTL
VILAI SPVLGSLAIWAKILSSFTDELLAYAKAGAAVEVLAARTVIAFGKKEL
ERYKNLEAKGIGIKAITANISIGAAFLIYASVALAFWCTSLVLSSEYSGIGVL
TVFVLIGAFSGQASPSIEAFANARGAAVEIFKIIDNKPSIDSTSKGKPDNLIK
NLFEKNVHFSYPSRKEVKILKLNKLVQSGQTVLVGNSSCGSKSTTVQLMQRLYDPTD
GMVICDGIJRTINVRHLREITGVVSOEPLVFATTAENIRYGRNTVMDIEKAVKE
SEAYVQVLDKARKRTTIVIAHRLSTVRNADVIAGFDGVIYKGNHDELMKEGITY
ANAVVQALDKARKRTTIVIAHRLSTVRNADVIAGFDGVIYKGNHDELMKEGITY
FKLVTMOTRNEIELENTESKESDALESPKDSGLIKRRSTRSTHAPQGDGR
KLCTKEDLENVPVPSWRLIKLNSTWEPYFVVGIFCAIINGLQPAFSIIFSRIGI
FTRDEDETRKQNSMFSVLVLGIISFTFFLOGTFCKAGEILTCKRYVWFRSM
LRQVSWFDDPKTITGALTTRLANDAAQVKAIGSRILAVITQNIANGTGIISLLYG
WQTLALLLAIPIIAIAGVEMKMLSGQALDKKLEGAKEIAEALENPRTVSLTR
EQKPEYMAQSLQVYRNSLRKAHIFGVFSITQAMMYFAGCFRFGALVANEFMN
FQDVLVFSALVFAGMAVGQVSSFPADYAKAVSAAHVIMIIEKSLIDSYSPHGLKP
NLBEGNTFNEVFNPTRPDIPVLGLSLEVKKGQTLALVSGSGCKSTVQVQLRFF
YDLPLAGSVLIDGKEIKHLNQWLRAHGIYSOBLPFDSCIAENIAYGDNRSVSHRE
IMQAAKANIHHTIETLPEKYNTPVKGKGTOLSGOKQRIAIARALVROPHILLDEA
TSALDTSEKVVQEAQDKAREGRTCTIVIAHRLSTQNIADLIIVFQNGKVAEHGTHQOL
LAQKGIYFSVMVQAGAKR"

BASE COUNT 1295 a 833 c 1008 g 1143 t

Alignment Scores:
Pred. No.: 1.66e-154 Length: 4279
Score: 1808.00 Matches: 363
Percent Similarity: 99.46% Conservative: 3
Best Local Similarity: 98.64% Mismatches: 2
Query Match: 98.47% Indels: 0
DB: 6 Gaps: 0

US-09-672-725C-6 (1-368) x AX105078 (1-4279)

QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValAlaGlnValAlaLeuAspLysAlaArg 20
Db 1694 ACCTCAGCTCTGCACACTGAAAGTGAAGCAGTGGTTCAGTGGCCCTGGATAAGGCCAGA 1753
QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1754 AAAGCGCGGACTACCATTTGTGATAGCTCATCTGTTTCTTACAGTTCTGTAATGCCGATGTC 1813
QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1814 ATTGCTGGTTTGTATGATGGAGTTCATTGTGGAGAAAGGAAATCATGCAACTCATGAAA 1873

```
QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
|||||
Db 1874 GAGAGGGCATTACTTCAAACTGTGCACATGCAGACAGAGAAATGAATTCAGTTA 1933

QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
|||||
Db 1934 GAAATGCCACTGGTGAATCCAAAAGTGAAAGTGATGCTTGGAAATGCTCTCCAAAAGAT 1993

QY 101 SerGlySerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
|||||
Db 1994 TCAGGTCACATTTAATAAAAAGAGATCAACTCGCAGGAGTATACATGCACCAAGGC 2053

QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
|||||
Db 2054 CAAGACAGAAAGCTTGGTACAAAAGAGACTTGAATGAGATGTACCTCCAGTTCTCTC 2113

QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
|||||
Db 2114 TGGAGGATTCGAAGCTCAACTCAACTGAATGGCTTATTGTGTGGTGTATATTTGT 2173

QY 161 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
|||||
Db 2174 GCTATTATAACGAGGCGCTGCACACAGCATTTTCAATAATATTTTCAAGGATTATAGG 2233

QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
|||||
Db 2234 ATCTTTACCCGAGATGAGGATCCTGAAACAAACAGCAGAAATAGTAACATGTTTCTGTA 2293

QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
|||||
Db 2294 TTGTTCTAGTCCCTGGAATATTTCTTTTATTACATTTTCTCTCCAGGCGTTCACATTT 2353

QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
|||||
Db 2354 GGCAAGCTGGGAGATCCTCACTAAGCGCTTCGATACATGTTTTCAGATCCATCGTG 2413

QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
|||||
Db 2414 AGACAGATGTCAGCTGGTTGATGACCTTAAAAACACCACCTGGAGCATTTGACAACAGG 2473

QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
|||||
Db 2474 CTGGCCAAATGATCGGCTCAAGTTAAAGGGGCTATAGTTTCCAGGCTTGCTGTCAATACC 2533

QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
|||||
Db 2534 CAGAATATAGCAAACTCTGGCAGAGCATTTATATATCTTATCTATCTGTTGGCAATTA 2593

QY 301 ThrLeuLeuLeuLeuAlaIleValProIleAlaIleAlaGlyValValGluMetLys 320
|||||
Db 2594 ACATTTTACTCTTAGCAATTGTACCATCATTTGCAATAGCAGAGTTGTTGAAATGAA 2653

QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
|||||
Db 2654 ATGTTGCTCGACAGACACTGAAGATAGAAAGACTAGAGGAGCTGGGAAGATGCT 2713

QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluLysPheGlu 360
|||||
Db 2714 ACAGAAGCCATCGAAAACCTCCGAACCTGTTGTTCTTGTACTCGGAGCAGAGTTTGA 2773

QY 361 HisMetTyrAlaGlnSerLeuGln 368
|||||
Db 2774 TACATGTATGCACAGAGTTTGCAG 2797

RESULT 7
AX105080
LOCUS AX105080 4279 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 24 from Patent WO0123540.
ACCESSION AX105080
VERSION AX105080.1 GI:13921230
KEYWORDS
SOURCE dog.
ORGANISM Canis familiaris
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 4279)
Stocker,P.J., Steimel-Crespi,D.T., Crespi,C.L., Reif,T.C. and
Patten,C.J.
P-glycoproteins and uses thereof
Patent: WO 0123540-A 24 05-APR-2001;
CENTEST CORPORATION (US)
FEATURES
    source
        1..4279
            Location/Qualifiers
                /organism="Canis familiaris"
                /db_xref="taxon:9615"
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAC37722.1"
                /db_xref="GI:13921231"
                /translation="MDPGRGRGSAEKNFKMGKKKKKKKKKKPTVSTFAMFRYSN
WIDRLYMLVGTMAAIIHGAAALPLMLVFGNWTDSFANAGISRNKTFVILINESITNNT
OHFIIHLEEMTYAYYSGIGAGVLVAAITQVSPWCLAAGRQILKIKQFFHAIHQ
EIGWFDVHDVGEINTRLDDVSKINEGDKIGMFFQFQSIATFFTFGIVFGTGGKMLT
VILASPVLGUSAAIWKAILSSFTDKELLAYAKAGAAVEVLAARTVIAFGGQKEL
ERYKNLEEKAGIGIKAITANSIGAAFLIYASYALAFWYGTSLVSSYSYQCVL
TVFFSVLIGAFSIGOASPIEAFANARGAAVEIKIIDNKPISDSYSGSHKPDNIKG
NLEFNHVSFSPRKEVKILGLNLKVQSGOTVALVNSGCGKSTTVQLMORLYDPTD
CMVICDQDIRTINVRHLREITGVVSOEPLFATIAENIRYGRNTVMDIEKAVKE
ANAYDFIMKLPNKFTDLVGERGAQLSGGQKORIAIALVRNPKILLDLDEATSLDTE
SEAVVADKARKGRITTVIAHRLSTVRNADVIAGDDGVIVEKHNHDELMKKGIV
FKLVTMOTRNEIELENTGESKESDALEMSPKDSLSLKRSTRSIIHAPQODR
KLGTEDLNEINPVPSFWRILKLSSTEMPYFVGIFCAIINGQLQAFSIIFFSRIIGI
PTRDEBTPKQNSNMFSLVLGLIISFTIFLQGTFFGKAGRILNKRLLRMVFRSM
LRDQVSWFDPKNTTGATTLRLANDAGVKGAIGSLRAVITQNLGTLGIIISLIG
WQTLILLAIPIITAIGVEMKMLSGQALADKKELEGAGIATEALENFTVIVSVLNR
EQDFYMTAQSLOVPYRNSLRKAHIFGVFSITQAMTYFSYAGCFRGAYLVANEFMN
FQDVLIVLSAIVFGMAVQSVSPADYAKAKVAHAHYIMIEKSPIDTSYSHGLRP
NTLEGNVTNEVNFYTRPDIPIVLQGLSLEVKKGQTLALVSGSGCGKSTVQVLLERF
YDPLAGSLVIDGKEIKHLNVQMLRAHLGIVSOEPLFDCSTAEINAYDNRSVVSHKE
IMOAKEANIHFETLPEKYTRVGDGKTLGSGQKORIAIALVRPHILLILDEA
TSALDTESEKVVQEAALDKAREGRTCIVIAHRLSTIQNLADLIIVFQNGKVEHGHQQL
LAQGDIFYSVMVQVQAKR"
BASE COUNT 1296 a 833 c 1008 g 1142 t
ORIGIN

Alignment Scores:
Pred. No.: 1,66e-154 Length: 4279
Score: 1808.00 Matches: 363
Percent Similarity: 99.46% Conservative: 3
Best Local Similarity: 98.64% Mismatches: 2
Query Match: 98.47% Indels: 0
DB: Gaps: 0

US-09-672-725C-6 (1-368) x AX105080 (1-4279)
QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
|||||
Db 1694 ACGTCAGCTCTGGACACTGAAAGTGAAGCAGTGTTCAGGTGGCCCTGGATAAGGCCAGA 1753

QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
|||||
Db 1754 AAAGGCCGGACTACCATTGTGATAGCTCATCGTTTGTCTACAGTTCTGTAATGCCGATGTC 1813

QY 41 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
|||||
Db 1814 ATTGCTGTTTGTGATGGAGTCACTTCGGAGAAGAAATCATGATCACTCATGAAA 1873

QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
|||||
Db 1874 GAGAAGGGCATTACTTCAAACTTGTCAACATGCAGACAGAGGAAATGAAATTCAGTTA 1933

QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
|||||
Db 1934 GAAATGCCCTGAGTGAATCCAAAAGTGAAAGTGATGCTTGGAAATGCTCTCCAAAAGAT 1993
```



```
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
Db 2114 TGGAGGATTCTGAAGCTGAAGCTCAACTGAATGGCCCTATTATTGGTGGTATATTTGT 2173
QY 161 AlalleIleAsnGlyGlyLeuGlnProIaPheSerIleIlePheSerArgIleIleGly 180
Db 2174 GCTATTATAAAGCGAGGCGCTGCAACAGCATTTTCAATAAATTTTCAAGGATTATAGGG 2233
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2234 AFTCTTACCAGATAGAGATCCTCGAAACAAACACAGAGATAGTAAATGTTTCTGTA 2293
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 2294 TTGTTTCTAGTCTCTTGGAAATATTCTTTTATTACATTTTCCCTCCAGGGCTTCACATTT 2353
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2354 GGCAAGCTGGGAGATCCTCACTAAGCGGCTTCGATACATGGTTTTCAGATCCATGCTG 2413
QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 2414 AGACAGGATGTCAGCTGGTTTGATGACCTTAAACACCATCGAGCATTCACAAACAGG 2473
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2474 CTTGCCAATGATCGCGCTCAAGTTAAAGGGCTATAGTTCAGGCTTGCTGCTATACC 2533
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 2534 CAGAATATAGCAAACTTGGGACAGGCATTATTATATCTTAATCTATGTTGGCAATTA 2593
QY 301 ThrLeuLeuLeuLeuAlaIleValProfileAlaIleAlaGlyValValGluMetLys 320
Db 2594 ACATTTTACTCTTAGCAATTTGTACCCATCATTTGCAATAGCAGGATGTTTGAATGAAA 2653
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 2654 ATGTTGCTGACAAAGCACTGAAAGATAAGAAAGACAGTACAGGAGCTGGGAAGATTGCT 2713
QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
Db 2714 ACAGAAGCATCGAAAACTTCCGAACGTGTTTCTTTGACTCGGGACGACAGATTGAA 2773
QY 361 HisMetTyrAlaGlnSerLeuGln 368
Db 2774 TACATGATGCACAGATTGCAA 2797
RESULT 9
LOCUS AF045016 4317 bp mRNA linear MAM 07-FEB-1998
DEFINITION Canis familiaris multidrug resistance p-glycoprotein (MDR1) mRNA,
complete cds.
ACCESSION AF045016
VERSION AF045016.1 GI:2852440
KEYWORDS .
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 4317)
Puel,O., Lepage,J.F., Alvinerie,M., Galtier,P. and Pineau,T.
Direct Submission
Submitted (28-JAN-1998) Pharmacology, INRA, BP 3, 180 Chemin de
Tournepieu, Toulouse Cedex 9 31931, France
Location/Qualifiers
source 1. .4317
/organism="Canis familiaris"
/db_xref="taxon:9615"
/tissue_type="colon"
/notes="beagle"
1. .4317
/gene="MDR1"
```

```
CDS
70..3912
/gene="MDR1"
/codon_start=1
/product="multidrug resistance p-glycoprotein"
/protein_id="AAC02113.1"
/db_xref="GI:2852441"
/translation="MDPEGRKGSAEKNFWKMGKSKKKEKKKPTVSTFAMFRYSNW
LDRLMLVGTMAAIHGAALPLMLVGNMTDSFANAGISRNKTPPVIINESITNTQ
HFNLHEEMETTYAYYIGAGVLVAAYIQVFWCLAAAGRLIKRKOFFHAIHQE
IGWFDVHDVDELNRLTDVSKINEGIDKGVGFQSIATFTFTGTVGTPGKWLTV
ILASIPVLGLSAAIWAKILSSFTKELLAYAKAGAAVEVLAAIRTVIAFGQKLE
RYNKEAKRIGIKKAITANISIGAFLLIYASVALAFWYGTSLVLSLEYTIGOVLT
VPSVLIGAFSTGQASPSLEAFANARGAAAEYFKIIDNKPDSIDSVSKSGHKPDNKG
LEFNVHFSYPSRKEKILKLNKAVQSGQTVALVNSCGGKSTVQLMQRLYDPTDG
MVCIDQIRTNVRLREITGVVSOEPLFATTAENIRYGRNVTWDEIEKAVKEA
NAYDFIMKLNKFDTLVGERGARLGGQKQRIARALVNPKEIILLDEATSDLTES
EAVVVALDKARKGRTTIVIAHRLSTVRNADVIAGDDGVIVKEGNHDELMKEGIYF
KLVTMOTRGNEIENATCESKESDLEMSPKDSGLIKRSTRRSHAPQGGDRK
LGTKEDELNENPVSVPWRILKLNSTENPVFVVGIFCAIINGGLQAFSTIFSRIGIF
TRDEPTEKRONSNFSLFLVLGISFTFFLQGTFTGKAGEILTKRLRYMFRSML
RODSWEDDPKNTGTALTTRLANDAAVKAGSLAVITQNIALGTGIIISLIYWG
OUTLLLAIVPIIATAGVVMKLSQALKDKKELEGAKIATEAENFRVTVSFTRE
QKFEYMAQSLQVPYRNSLRKAHIFCSFISITQAMMYFSYAGCFRGLYLVANEPNF
QVLLVFSIVFEGAMAVGOVSFPADYAKAVSAHVIMIIEKSLIDSYSHPGLKPN
TLEGWTFNEVFNPTREDIPVLOGLSLEVKKGTALVSGSGGCKSTVSVVQLLERY
DPLAGSVLDGKEIKHLNVOMLRAHLGIVSQEPILFDCSIAENIAGNSRVSHHEEY
MQAAKEANIHQHFIEITPEKYNTRVGDKGTQSGGQKQRIATARALVROPHILLDEAT
SALDTESEKVVQELDKARKREGRTCVIAHRLSTIQNIADLIIVVFQNGKVKHEGTHQOLL
AOKGIYFSMISVQAGAKR"
BASE COUNT 1293 a 844 c 1019 g 1161 t
ORIGIN
Alignment Scores:
Pred. No.: 8,97e-154 Length: 4317
Score: 1800.00 Matches: 362
Percent Similarity: 99.18% Conservative: 3
Best Local Similarity: 98.37% Mismatches: 3
Query Match: 98.04% Indels: 0
DB: 4
US-09-672-725c-6 (1-368) x AF045016 (1-4317)
QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArg 20
Db 1744 ACCTCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGTGGCCCTGGATGAAGCCAGA 1803
QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1804 AAAGCGCGGACTACCATTTGTGATGCTCATCTGTTCTCTACAGTTCGTATGCCGATGTC 1863
QY 41 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1864 ATTGCTGGTTTGTATGATGAGTCAATTGTGGAGAAAGAAATCATGATGAATCATGAAA 1923
QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluLeuAspLeu 80
Db 1924 GAGAAGGCATTTACTTCAAACTTGTCACATGCAGACAAGAGAAATGAAATGAGTTA 1983
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 1984 GAAATGCCACTGGTGAATCCAAAAGTGAAGTGAATGATGCTTGGAAATGCTCCAAAAGAT 2043
QY 101 SerGlySerSerLeuIleLysArgSerThrArgSerIleHisAlaProGlnGly 120
Db 2044 TCAGGGTCCAGTTTATATAAAAAAGAGATCAACTCGCAGGAGTATACATGCACCAACAGGC 2103
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
Db 2104 CAAGACAGAAAGCTTGGTACAAAAGAGGACTTGAATGAGATGTTCCCTTCAGTTCCTTC 2163
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
Db 2164 TGGAGGATTCTGAAGCTGAACCTCAACTGAATGCCCTTATTGTTGTTGGTATATTTGT 2223
```


QY 161 AlaIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 2224 GCTATTATAACGGAGCGCTGCACACGATTTTCAATAATATTTTCAAGGATTTATAGG 2283
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2284 ATCTTTACCGAGATGAGGATCTCTGAACAAACGACAGATAGTAACATGTTTCTGTA 2343
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 2344 TTGTTTCTAGCTCTCGAATATTTCTTTTATATACATTTTCTCAGGCGCTTCACATTT 2403
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2404 GGCAAGCTGGGAGATCCCTACTAAGCGCTTCGATACATGTTTTCAGATCCATGCTG 2463
QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 2464 AGACAGATGTCAGCTGGTTGATGACCCCTAATAACACACCTGGAGCATTTGACACAGG 2523
QY 261 LeuAlaAsnAspAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2524 CTTGCCAATGATCGGCTCAAGTTAAGGGCTATAGTTCCAGGCTTCTGTCATTACC 2583
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 2584 CAGATATACCAATCTTGGACAGCATATATATCTTAATCTATGTTGGCAATTA 2643
QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
Db 2644 ACATTTTACTCTAGCAATTTGACCATCATTTGCAATAGCAGGAGTTGTTGAAATGAAA 2703
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 2704 ATGTGTCTGGCAAGCACCTGAAAGATAGAAAGAGCTAGAAGAGCTGGGAAGATTGCT 2763
QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
Db 2764 ACAGAAGCATCGAAACTTCGAACTGTGTTCTTTGACTCGGAGCAGAGAAGTTTGAA 2823
QY 361 HisMetTyrAlaGlnSerLeuGln 368
Db 2824 TACATGATGCACAGATTTGCAA 2847
RESULT 10
AX105059
LOCUS AX105059 4317 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 3 from Patent W00123540.
ACCESSION AX105059
VERSION AX105059.1 GI:13921211
KEYWORDS
SOURCE
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 4317)
Stockert,P.J., Steimel-Crespi,D.T., Crespi,C.L., Reif,T.C. and Patten,C.J.
TITLE P-glycoproteins and uses thereof
JOURNAL Patent: WO 0123540-A 3 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
source Location/Qualifiers
1. .4317
/organism="Canis familiaris"
/db_xref="taxon:9615"
70. .3912
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37719.1"
/db_xref="GI:13921212"
/translations="MDPEGRKGSAEKFWKSKKSKKKEKKKPTVSTFAMFRYSNW
LDRLYMLVGTWAAIIHGAALPLMLVFGNWTDSFANAGISRNKTFPVIINESITNNTQ

BASE COUNT 1293 a 844 c 1019 g 1161 t
ORIGIN

Alignment Scores:
Pred. No.: 8.97e-154 Length: 4317
Score: 1800.00 Matches: 362
Percent Similarity: 99.18% Conservat: 3
Best Local Similarity: 98.37% Mismatches: 3
Query Match: 98.04% Indels: 0
DB: 6 Gaps: 0

US-09-672-725C-6 (1-368) x AX105059 (1-4317)

QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
Db 1744 ACCTCAGCTCGACACCTGAAAGTGAAGCTGAGTGGTTCAGTGGCCCTGGATAAGGCCAGA 1803
QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1804 AAGGCCGAGCTACCATTTGTATAGCTCATCTGTTGTCTACAGTTCTGTAATGCCGATGTC 1863
QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1864 ATTGCTGGTTTGTATGATGGAGTCAATGTGGAGAAAGGAATCATGATGACATCATGAAA 1923
QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
Db 1924 GAGAGGGCATTTTACTTCAAACTTGTCAATGCACAGACAGAGGAATGAAATTCAGTTA 1983
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 1984 GAAATGCCACTGGTGAATCCAAAAGTGAAGTGAATGAGTGTTCCTTCAGTTTCTCTTC 2043
QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
Db 2044 TCAGGTCGAGTTTATAAAAAAGATCACTCCAGAGTATACATGCACACAGGC 2103
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
Db 2104 CAAGACAGAAAGCTTGGTACAAAAGAGGACTTGAATGAGATGTTCTTCAGTTTCTCTTC 2163
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCys 160
Db 2164 TGGAGGATTCGAGCTGAACCTCAACTGAATGGCCTTAATTTGTGGTATATATTTGT 2223
QY 161 AlaIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 2224 GCTATTATAACGGAGCGCTGCACACGATTTTCAATAATATTTTCAAGGATTTATAGG 2283
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2284 ATCTTTACCGAGATGAGGATCTCTGAACAAACGACAGATAGTAACATGTTTCTGTA 2343

HFNHLEEMTYAYYYGIGAGVLAAAYIQVSWCLAAAGRQLKTRKQFFHAIMRQE
IGWFDVHDVGLNRLTDDVSKINEGDKVGMFFQSIAFTFTGTVGTPGKKTLV
ILASPVLGSAAIWAKILSFTDKELLAYAKAGAAVEELAAIRTVIAFGQCKELE
RYNKELEAKRIGIKKATIANISIGAFTLLIYASALAFWYGTSLVSSEYITGQVLT
VFTSVLIGAFSIGQASPSIEAFANARGAAVEIPKIDNKPSIDYSKSGHKPDNKG
LEPKNVHSPSRKEKILKLNKLVQSGOTVALVNSCGGKSTTVQLMORLYDPTDG
MVCIDQODIRTINVRHLREITGVVQSEPVLFATIAENIRYGRNVTMDIEKAYKEA
NAYDFIMKLPNKEDTLVGERGARKSGOKORIAIALVRNPKILLDELMSALDTES
EAYVOVALDKARKGRTTVIAHRLSTVRNADVIAGDFDGVIVEKHNHDELMKPKYIF
KLVMTOTRGNEIELENTAGSKESDALEPKSDGSSLIKRSTSRSHAPQODRK
LGTCKEDLNENVPVSPWRLLKLNSTWEPYVVGICAIINGGLOPAFSTIFSRIGIF
TRODDPDKTQNSNMFVLEVLGILISFTFFLOGFTFGKAGEILTKRLKRVYFRSM
QVDFPDDPKNTALTTRLANDAAQVGAISRLAVIQNTIANLGTGILISLIYGM
QLTILLAIYPIIAIAGVEMKLSQALKKELGAGKIAATEALENRTVVVSLRE
OKFMYAQSLQVPIYRNSLRKAHIFGVSFISITQAMMYFSTAGCFRGAALVANEFNF
QDLVLFSAIVFGAMVQSVFAPDYAKAKVSAARVIMIEKSPIDISYVQGLKPN
TLBNWTFNEVFNYPTRPDI PVLGLSLEVBKGGOTLALVSSGCCSKSTVQLLERFY
DPLAGSVLDGCKELKHLNWLRAHLGINSOEPILPDCSTAEANIAGXDNRSVYSHEE
MORAKEANIHF IETLPKRYNTRVGDGTQLSGGOKORIAIALVRNPKILLDELMS
SALDTESKIVQEALDKAREGRTCVIAHRLSTQNLADLLIVFQNGKVKHBGHTHQLL
AQKIYFSMISVQAGKR"


```
Qy 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
|||||
Db 2344 TTGTTTCTAGTCCCTGGAAATATTTCTTTATTACATTTTCTCCAGGGCTTCACATTT 2403
Qy 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
|||||
Db 2404 GGCAAGCTGGGGAGATCCTCCTCAAGGGCTTCGATACATGTTTTCAGATCCATGCTG 2463
Qy 241 ArgGlnAspValSerTTPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
|||||
Db 2464 AGACAGATGTCAGCTGTTGATGACCCCTAAACACCACTGGAGCATTTGACAACAGG 2523
Qy 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
|||||
Db 2524 CTTGCCAATGATGGCTCAAGTTAAAGGGCTATATGTTCCAGGCTGCTGTCATACC 2583
Qy 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
|||||
Db 2584 CAGAAATATAGCAATCTTGGGACAGCATTTATATATCTTAATCTATGTTGGCAATTA 2643
Qy 301 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
|||||
Db 2644 ACATTTTACTTCTTAGCAATGTACCCATCATTGCAATAGCAGGAGTTGTGAATGAA 2703
Qy 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
|||||
Db 2704 ATGTTGTCTGGACAAGCACTGAAGATAAGAAAGAGCTAGAAAGAGCTGGGAAGATGCT 2763
Qy 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGlnLysPheGlu 360
|||||
Db 2764 ACAGAAGCCATCGAAAACCTTCCGAACCTGTGTTTCTTGTACTCGGGAGCAGAAAGTTGAA 2823
Qy 361 HisMetTyrAlaGlnSerLeuGln 368
|||||
Db 2824 TACATGATGCACAGATTGTCAA 2847

RESULT 11
AB066299
LOCUS
DEFINITION
Canis familiaris MDR1 mRNA for multidrug resistance p-glycoprotein,
partial cds.
ACCESSION
AB066299
VERSION
1
KEYWORDS
GI:18253110
SOURCE
Canis familiaris lymphoma cell_line:GL-DOX4000 cDNA to mRNA.
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
1 (sites)
Dozumi,K. and Nakaichi,M.
Canine MDR
Unpublished
2 (bases 1 to 2005)
Dozumi,K. and Nakaichi,M.
Direct Submission
Submitted (19-JUL-2001) Kayo Uozumi, Yamaguchi Univ., Department of
Veterinary Surgery; 1677-1 Yoshida, Yamaguchi, Yamaguchi 753-8515,
Japan (E-mail:miniturejinpinsher@yahoo.co.jp, Tel:81-83-933-5898,
Fax:81-83-933-5898)
FEATURES
Location/Qualifiers
1..2005
/organism="Canis familiaris"
/db_xref="taxon:9615"
/cell_line="GL-DOX4000"
/cell_type="lymphoma"
1..2005
/feature="MDR1"
<1..>2005
/feature="MDR1"
/codon_start=1
/product="multidrug resistance p-glycoprotein"
/protein_id="BAB83959.1"

gene
CDS
```

```
/db_xref="GI:18253111"
/translation="AGAAVEEVLAAIRTVIAFGGQKLEERYNNKLEPAKRIGIKKAI
TANISGAFLIYASYALAFWISGYSLSSEYSIGVLTFFSVLIGAFSIGQASPS
IEAFANAGAAEYIEFIIDNKPSIDYSKSGHKPDNIKNLEFKNHVSYSFSEVKI
LKGLNKVQSQTVALVGNSSGCKSTVOLMRLYDPTDGMVCDIGDQIRINVRHLR
ETGVVSPQEVFLFATTIENIRYGRNVTDMEIEKAVEANAYDFIMKLPNKFDTLVG
BGRALSGQKORIAIALALVRNPKILLDEATSDALDESEAVQVADLKARKGRTTI
VTARLSIVRNADVATGPDGVIVEKGNHDELMKKGIIYFKLVTMOTRGNEIELENAT
GESKSEDALEMSPKDSSSLIKRRSTRSRSHAPQOQKRLGTDKEDLNEVPVSFWR
ILKLNSTWPFYFVFCALINGLOPAPSIIFSRIGITFTRDEDEPTKRONNMFVS
LEFLVIGIISFTFFLQGTGKAGEILTKRLRYMVRSMRLQDVSFEDPKNTTGALT
TRLANDAAGKVAIGSRILAVITONIANLGTGIIISLYGWOLTLLLIIVPIIAGV
VBMKLSGQALKDKKELSGAGKIATEATENFTVVSILTREQKE"
BASE COUNT 610 a 371 c 487 g 537 t
ORIGIN
Alignment Scores:
Pred. No.: Length: 2005
Score: 1764.00 Matches: 355
Percent Similarity: 99.17% Conservative: 2
Best Local Similarity: 98.61% Mismatches: 3
Query Match: 96.08% Indels: 0
DB: 4 Gaps: 0
US-09-672-725C-6 (1-368) x AB066299 (1-2005)
Qy 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
|||||
Db 925 ACCTCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCTGGTAAGCCAGA 984
Qy 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
|||||
Db 985 AAAGCCGACGACATTCATGATGATCATCTGTTTGTCTACAGTTCCGTAATGCCATGTC 1044
Qy 41 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
|||||
Db 1045 ATTGCTGGTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1104
Qy 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluLeuAspLeu 80
|||||
Db 1105 GAGAAGGGCATTTACTTCAACTTGTCACTGACAGCAAGAGAAATGAAATGAGTTA 1164
Qy 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
|||||
Db 1165 GAAATGATCCACTGGTGAATCCAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1224
Qy 101 SerGlySerSerLeuLysArgSerThrArgArgSerIleHisAlaProGlnGly 120
|||||
Db 1225 TCAGGGTCCAGTTTAATAAAAAAGAGATCAACTGCAGGAGTATACATGCACCAAGGC 1284
Qy 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
|||||
Db 1285 CAAGACAGAAAGCTTGGTACAAAAGAGGACTTGAATGAGAATGTTCCCTTCAGTTTCCCTTC 1344
Qy 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
|||||
Db 1345 TGGAGGATTCGAACTGAACTCAACTCAATGAGCTTGAATGAGTGAATGAGTGAATGAGTGA 1404
Qy 161 AlalIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleGly 180
|||||
Db 1405 GCTATTATAAAGGGAGGCTGCAACCCAGCATTTTCAATAATATTTTCAAGGATTTATAGG 1464
Qy 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
|||||
Db 1465 ATCTTTACCCGAGATGAGGATCTCGAAACAAACAGACAGATAATGATGATGATGATGATG 1524
Qy 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
|||||
Db 1525 TTGTTTCTAGTCCCTGGAAATATTTCTTTTATTACATTTTCTCCAGGGCTTCACATTT 1584
Qy 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
|||||
Db 1585 GGCAAGCTGGGGAGATCCTCCTCAAGGGCTTCGATACATGTTTTCAGATCCATGCTG 1644
```

```
QY 241 ArgGlnAspValSerTrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
|||||
Db 1645 AGACAGGATGTGACGTGGTTGATGACCTTAAACACACACATGGAGCATTCACACACAGG 1704
|||||
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
|||||
Db 1705 CTTGCCAATGATGCGGCTCAAGTTAAAGGGCGTATAGGTTCCAGGCTTGCTGTCATTACC 1764
|||||
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleThrGlyTrpGlnLeu 300
|||||
Db 1765 CAGAATATAGCAAACTCTGGGACAGCATTTATATATCTTAATCTATGTTGGCAATTA 1824
|||||
QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
|||||
Db 1825 ACACITTTTACTCTTAGCAATTTGACCATCATGCAATGACAGGAGTTGTTGAATGAAA 1884
|||||
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysGlyLeuGluGlyAlaGlyLysIleAla 340
|||||
Db 1885 ATGTTGCTGGCAAGCAAGCTAGAAAGATAGAAAGAGCTAGAAAGAGCTGGGAAGATTGCT 1944
|||||
QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
|||||
Db 1945 ACAGAAAGCATCGAAACTTCGAACTGTTGTTCTTACTCGGGAGCAGAAAGTTTGAA 2004
|||||
RESULT 12
AB029153 3489 bp mRNA linear MAM 30-NOV-1999
LOCUS Felis catus multi-drug resistance related mRNA, partial cds.
DEFINITION AB029153
ACCESSION AB029153.1 GI:6472652
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Tsuji moto, H. and Okai, Y.
TITLE Molecular analysis of multidrug resistance in lymphoma cells in the
Cat
JOURNAL Am. J. Vet. Res. (1999) In press
REFERENCE 2 (bases 1 to 3489)
AUTHORS Tsuji moto, H. and Okai, Y.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1999) Hajime Tsuji moto, The University of Tokyo,
Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan
(E-mail: atsujihongo.ecc.u-tokyo.ac.jp, Tel. +81-3-5841-8004,
Fax: +81-3-5841-8178)
FEATURES
Location/Qualifiers
1..3489
/organism="Felis catus"
/db_xref="taxon:9685"
/cell_line="adriamycin-resistant subline of FT-1"
/cell_type="feline lymphoma cell line"
<1..>3489
/note="multi-drug resistance related mRNA"
/codon_start=1
/protein_id="BAA87071.1"
/db_xref="GI:6472653"
/translation="HGAALPLMLLVFGDMTOSFANAGISRNLTTLNITGESIVNDSYF
INRLEEMFTYAYVYSGIGAGVLAAYIQVSWCLAAGRQILKIRKQFHAIMROEVG
WFDVHDVGLNRLTLDVDSKINEGIDKIMFQSMATFFIGFVGFPRGKWLTVIL
ATSPVLGSLAAWAKLTSSFTKELLAVAKAGAAVEELAAIRTVIARFGGQKLELERY
NNRLERAKRGIKKAITANISIGIAFLLIYAYALAFWYGTSLVLSHEYSGOVLTVF
FVSLVIGAGVGGAQSPSIEAFANARGAAEYFIKIDNKEISIDSYSKNGKHPDNKGNLE
FRNVHPSYREKVLKGLNLKLVQSGGTVALVNSGCGSTTVOLMDPTDGMV
SIDQDRTINRYIVVYSGQEVLFATIAENIRYGRNVTMEIRKXAEANA
YDFIMKPLNGFTLVGERGAOLSGGQKQRIARALVRNPKILLDEATSDLSSEA
VQVADLKARKGRTTIVAHRLSTIRNADVIAGFDDGVIVKGNHDELMKEGIFYKL
VTMQTRGTELENAYVESISEIDALENSPKDSGSLIRRRSTRKSIHASQOQDRKLG
```

```
TKENLDERVPPVSWRILKLNITETHPYFVVGIFCAILINGLOPAESVILSRIGVETR
VDEPETKQNSNIFSLFLVLGLIISITFFLOGFTFGAGEILLTKLRIMWPRSMLRQ
DYSWDFDKNTTGTALTTRLANDAAQVKAIGSLAITONTIANLGTGIIISLIYQWL
TILLALLVPIIAAGVEMKSLGQALDKKELEGKATEAIEFNFTVSLTREQK
FEYWAQSLQVPYRNSLRKAHIFGIFSITQAMVFSYAGCFRFGAYLVAHFMDFOQ
VLIVFSAIYFGAMAYGVSSFPADYAKAKVSAHVIMIEKIPLDISVSTEGLMPTNL
EGNVTVEYMFNPTRPDIPVLOGLSLEVKKQKTLALVSGSSCGAGSTVQVLLERSYDP
MAGTVLIDGKEIKHLNVOLRAHMGIVSQEPILFDCSIGENIATGDSNRVVSQEEIVR
AAKEANIHPFIETLPDKYNTYRVGDKGTQLSGGQKQRIARALVRQPRILLDEATSA
LDTESKVVQOEALDKAR"
BASE COUNT 1015 a 716 c 847 g 911 t
ORIGIN
Alignment Scores:
Pred. No.: 8.4e-145 Length: 3489
Score: 1700.00 Matches: 339
Percent Similarity: 96.74% Conservative: 17
Best Local Similarity: 92.12% Mismatches: 12
Query Match: 92.59% Indels: 0
DB: Gaps: 4
US-09-672-725C-6 (1-368) x AB029153 (1-3489)
QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
|||||
Db 1495 ACATCAGCCTTGGACAGTGAAGTGAAGCAGTGGTTCAGGTGGCCCTGGATAGGCCAGA 1554
|||||
QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
|||||
Db 1555 AAAGCGCGGACCACCATAGTGGTAGCTCATCGTTGTCTACAATTCGTAATGCTGATGC 1614
|||||
QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
|||||
Db 1615 ATCCCGCGTTTGTATGATGGAGTCATTGTGGAGAAAGGAATCATGATGAACATCGAATA 1674
|||||
QY 61 GluLysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
|||||
Db 1675 GAGGAAGCATTTACTTCAAACTTGTCAATGTCAGACACAGAGGAATGAAATGAATTA 1734
|||||
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
|||||
Db 1735 GAAATGCGATTTATGAATCCATAAGTGAATTCATGCTTGGAAATGCTCCAAAAGAT 1794
|||||
QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
|||||
Db 1795 TCAGATCCAGTTTAAATAGAGAGAGATCAACCCGCAAGAGATATACATGATCACAAGGC 1854
|||||
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
|||||
Db 1855 CAAGATAGAAAGCTGGTACAAAAGAGAACTTGGATGAACGCTACCTCCAGTTTCCTTC 1914
|||||
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluThrProThrPheValValGlyIlePheCys 160
|||||
Db 1915 TGGAGGATTCGAAGCTCAATATAACTGAATGGCTTATTTTGGTGGTATATTCTGT 1974
|||||
QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
|||||
Db 1975 GCTATTATAAATGAGGCGCTGCAACCCAGCATTTTCAGTAATACTTTCAGAGATTATAGG 2034
|||||
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
|||||
Db 2035 GTTTTACCAGATTGAGGATCCTGAAACAAAACACAGAAATAGTAAATATTTTCATTA 2094
|||||
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
|||||
Db 2095 TTGTTTCTAGCTCTTGGAAATATTCTTCTATTACATTTTTCCTTCAGGGCTTCACGTTT 2154
|||||
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
|||||
Db 2155 GCACAGGCTGGGAGATCCTCACCAGCGGCTCCGATACATGGTCTTCAGATCCATGCTA 2214
|||||
QY 241 ArgGlnAspValSerTrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
|||||
```

```
Db 2215 AGACAGGATGTGAGCTGTTGATGACCTAAAAATACTACTGGAGCATTTAAACAACAGG 2274
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2275 CTTGCCAACGAGCTGCCCAAGTTAAAGGGCTATAGGTTCCAGGCTTGCTATCATACC 2334
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 2335 CAGAAATATAGCAATCTTGGACAGGCATTATATATCTTAATCTAATGCTGCATG 2394
QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
Db 2395 ACACCTCTTACTCTTAGCAATTTGATCCCATCTTGCATATGACGAGGAGTTGTTGAAATGAAA 2454
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 2455 ATGTTGCTGGACAGCACTGAAGATGAAGAACTAGAGGTGCTGGGAAGTTGCT 2514
QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluLysPheGlu 360
Db 2515 ACTGAAGCAATCGAAACTTCGGAACCTGTTGCTCTTTGACTCGAGAGCAAGCTTTGAA 2574
QY 361 HisMetTyrAlaGlnSerLeuGln 368
Db 2575 TATATGTATGACAGAGTTTCAG 2598

RESULT 13
AX108654
LOCUS
DEFINITION Sequence 1 from Patent WO0123565.
ACCESSION AX108654
VERSION AX108654.1 GI:13923886
KEYWORDS crab-eating macaque.
SOURCE Macaca fascicularis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Macaca.
1 (bases 1 to 4186)
Stocker,P.J., Steimel-Crespi,D.T. and Crespi,C.L.
P-glycoproteins from macaca fascicularis and uses thereof
PATENT: WO 0123565-A 1 05-APR-2001;
GENTEST CORPORATION (US)
FEATURES
source
1..4186
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
100..3942
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:13923887"
/translation="MDLEDGRNGGAEEKKNFFLNKSKDKKERKPTVSVFSMFRYSN
WLDKLYVVGTLAAIIGHAGLPLMLFGDMDTFANAGLGDGLALLTSNMTDITV
PVNLEEDTRVAYYSIGICAGVLAAYIQVSWFLAAGRQIHKIRKQPFHAIHQEI
GWNFDHVGELNRYLDDYSGKINEGIKGMFQSMATFTFTGIVFRGKWLIV
LATSPVLGSAAWAKILSSFTDKELLAYAKAGAAVEVLAIRVIAFGGKKELER
YKNLEEARIGIKKAITANSIGAAFLIYASALAFWIGTILVLSKRSIGQVLT
FFSLVLAGSVGASPISEAFANARAAFEIKIIDNKPSIDSYSKSGHKPDNIKGNL
EFNRFHSPSRKEVKILKLNKLVQSGQTVLVNSGCGSKSTVQLMQRLDPTEGM
VSDVGDITRTINVFREIIGVVSQEPVLFTAIENIRYGREYVMDIEKAVKAN
AYDFIMKLPQKFTLVGERGAOLSGGKORIAARALVRNPKILLDEATSLDTSE
AVQVQALDKRGRTTIVIAHRLSTVRNADVTAGPDDGVIVEKSHDELMEKGYIFK
LYTMQTAQNEILENAADESKSEIDTLEMSHDSGSLRKRKSTRSRVSGQOKKL
SKELADESIPVPSFWKIMKLNTEWYFVGVFCALINGLOPAPAVIFSIIIGFT
RNDAAETKRONLSFLSLVLGIVSIFTFQGTGFKAGEILTKRLRYFNRMLR
QVSWFDDPKNTGALTTRLANDAAQVKAIGSRILAIITONIANLGSTIIISLIYQW
LTLLILAIVPIIAIAGVEMKMLSGQALKDKKELEGAKIATEATENFTVVSLEQEF
KPEHMYDSIQVYPRNSLRKAHIFGTFSTQAMVFSVAGCFRGALVAHLSMFE
DVLVYSAVFGAMAVGQSSFPAPYAKAKVSAAHIMIEIKTILDSITSTGLAPNT
LGNVTFNVPVNPTRLDIPVLQGLSLVKKGTALVQSSGCKSTVQLLERFYD
PLAGKVLDDGKEIKOLNQMLRAHLGIVSQEPIILFDCSISENIAYGDSRVVSQEIV
RAAKEANIHAFTIESLPNKYSTRVGDKGTQLSGGQKORIAARALVRQPHILLDEATS
```

```
ALDTESEKVVQEALDKAREGRTCIIVIAHRLSTIQNADLIIVPQNGVRKHEGTHQOALLA
QKGIFFSVMYSVOAGAKRQ"
BASE COUNT 1226 a 799 c 1039 g 1122 t
ORIGIN
Alignment Scores:
Pred. No.: 1,06e-142 Length: 4186
Score: 1678.00 Matches: 333
Percent Similarity: 95.38% Conservative: 18
Best Local Similarity: 90.49% Mismatches: 17
Query Match: 91.39% Indels: 0
DB: 6 Gaps: 0
US-09-672-725C-6 (1-368) x AX108654 (1-4186)
QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
Db 1771 ACCTCAGCCCTGGACACAGAAAGTGAAGCAGTGGTTTCAGTGCTCTGGATAAGGCCAGA 1830
QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1831 AAAGGTCGGACCACTTGTGATAGCTCATCGTTTGTCTACGGTTCGTATATGCCGACGTC 1890
QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1891 ATCGCTGGTTTCGATGATGGAGTCATTTGGAGAAAGGAATCATGATGAGTCATGAAA 1950
QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
Db 1951 GAGAAAGGCATTTACTTCAAACTTGTCAACATGTCACAGCAGCAGGAAATCAAATTTGAATTA 2010
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 2011 GAAATATGACGCTGATGAATCCAAAGTGAATGATACCTTGGAAATGCTTTCACATGAT 2070
QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
Db 2071 TCAGGATCCAGCTCTAATAAGAAAAGATCCACTCGTAGGAGTGTCCGTGGATCACAGGC 2130
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 140
Db 2131 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGTATACCTCCAGTTTCCTTT 2190
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValGlyIlePheCys 160
Db 2191 TGGAGGATATGAAAGCTAAATTTAACTGAGTGGCCCTATTTTGTGTGGTGTATTTGT 2250
QY 161 AlaIleAlaAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 2251 GCCATTATAATGGAGGCTCGCAACACGAGCATTTGCAGTAATATTTTCAAAGATTATAGG 2310
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2311 ATTTTACAAGAAATGATGATCGGAAACAAACAGACAGAAATAGTAACCTGTTTTCACATA 2370
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 2371 TTGTTTCTAGTCTTGGAAATGTTTCTTTATTACATTTTCTCTTCAGGGCTTCACATTT 2430
QY 221 GlyLysAlaGlyIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2431 GGCAAAGCTGGAGAGATCTCCACCAAGCGCTCCGATACATGATGTTTCCGATCCATGCTC 2490
QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 2491 AGACAGGATGTGAGCTGGTTTTCATGCCCTAAACACACACCTGGAGCATTTGACTACCAGG 2550
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2551 CTCGCAATATGATGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCATATAATTACC 2610
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
```

Db 2611 CAGAAATAGCAAAATCTTGGGACAGGAATAATATATATCTTAATCTATGGTTGGCAACTG 2670

Qy 301 ThrLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320

Db 2671 ACACCTGTACTCTTACCAATGTACCCATCATTCGAATAGCAGGAGTGTGTAATGAAA 2730

Qy 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340

Db 2731 ATGTTGCTGGACAAGCACTGAAGATAGAAGAACTAGAGGTGCTGGGAAGATCGCT 2790

Qy 341 ThrGluAlaIleGluAsnPheThrValValSerLeuThrGlnGluGlnLysPheGlu 360

Db 2791 ACTGAAGCAATAGAAACTTCGAACTGTGTTCTTTGACTCAGGACGAGAAGTTTGAA 2850

Qy 361 HisMetTyrAlaGlnSerLeuGln 368

Db 2851 CATATGTATGATCAGAGTTTGAG 2874

RESULT 14

AX108656

LOCUS

DEFINITION

Sequence 3 from Patent WO0123565.

ACCESSION

AX108656

VERSION

AX108656.1 GI:13923888

KEYWORDS

crab-eating macaque.

SOURCE

Macaca fascicularis

ORGANISM

Eukaryota; Chordata; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae; Cercopitheciinae; Macaca.

REFERENCE

1 (bases 1 to 4195)

AUTHORS

Stocker,P.J., Steimel-Crespi,D.T. and Crespi,C.L.

TITLE

P-glycoproteins from macaca fascicularis and uses thereof

JOURNAL

Patent: WO 0123565-A 3 05-APR-2001;

GENEST

CORPORATION (US)

FEATURES

source

1..4195

/organism="Macaca fascicularis"

/db_xref="taxon:9541"

100..3951

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC37765.1"

/db_xref="GI:13923888"

/translation="MDLEGDRNGGAKEKNFKLNKSKDKKERKPTVSVFMSFRYSN WLDKLYMVVGTLLAIIHGAGLPLMLLVFGDMTDFANAGNLGDGLALLFNNTSSNIT DTVPVNLDEDMTRYVYYSIGIGAGLVAAVIOVSFWCLAAAGROIHKIRKQFFHAIHR QEIGFQVDVHGEINLRLTDVSKINEGIGDKGMFFQSMATFTFTGIVGTRGWKLT LVILAI5PVLGSLAAWAKILSSFTDKRELLAYAKAGAAVEVLAAITVIAFGQKKE LERVNKLDEAKRIGIKKAITANISIGAFLIITASTALAFWYGTITVLVLSKEYSIGV LTVF5VLIGAF5GQASPSKEAFANARGAAFEIFKIIDNKP5ID5YSK5GHKPDNIK GNLEFRNVHFSYPSRKEVKILKGLNVQSGQTVLVN5GCGSK5TTVLMQRLYDPT EGMVSDQDITRNVRLEIIGV5SOEPLVFIATIAENIRYGREVDVWDEIEKAVK EANA5DETMKLPQKDTLVGRGAOLSGGOKORTIARALVRNPKILLDEAT5ALDT E5EAVQVQALDKARKGRTTIVIAHLS5TVRNADV5AGFDG5IVEKGNHDLMEKGI YfKLV5TAGNE5LENADE5SEIDTLEM5SHD5GSS5LIRK5TR5R5VR5GQGD RLK5TEALDE5IP5P5WR5ML5KLN5L5T5W5Y5F5V5F5CA5L5ING5L5Q5L5AF5V5K5I5G I5FR5ND5DA5TR5Q5N5F5L5L5V5L5G5I5V5F5F5L5Q5F5TG5AGE5ILT5K5RL5RV5FR5 M5LR5Q5V5F5D5P5K5T5AT5ALT5RL5AND5AQ5V5KA5G5S5RL5AI5T5IAN5L5G5I5I5LY GW5OLF5L5L5V5PI5T5I5AG5V5M5K5G5O5L5K5D5K5E5G5AK5I5A5E5I5EN5FR5V5S5LT Q5O5K5E5H5M5D5Q5L5Q5V5P5N5L5K5AH5IG5ITE5FT5Q5AM5F5Y5AC5FR5G5L5V5AH5SLM SP5ED5L5V5FS5AV5G5MA5V5Q5V5SF5P5DY5AK5V5SA5HI5M5I5E5K5T5L5D5S5Y5E5G5L P5N5L5E5G5N5T5F5E5V5N5F5Q5I5D5I5P5V5Q5L5E5V5K5Q5T5L5AL5V5G5G5G5K5T5V5Q5LLER FYD5PL5AG5V5L5D5K5E5I5K5Q5N5O5W5L5R5H5G5I5V5Q5E5P5L5F5D5C5I5S5EN5I5AY5G5N5R5V5S5E ET5V5RA5K5AN5L5H5AT5E5I5P5N5K5Y5TR5V5G5D5K5Q5L5G5G5Q5R5I5A5I5AR5L5P5H5ILL5DE AT5AL5D5TES5K5V5Q5E5L5DK5ARE5G5RT5C5I5V5I5A5H5RL5T5I5Q5N5AD5L5I5V5F5Q5N5R5K5E5H5G5THQQ L5L5A5O5K5I5F5S5M5V5Q5A5K5R5"

BASE COUNT 1230 a 801 c 1039 g 1125 t

ORIGIN

Alignment Scores:

Pred. No.: 1,06e-142 Length: 4195

Score: 1678.00 Matches: 333

Percent Similarity: 95.38% Conservative: 18

Best Local Similarity: 90.49% Mismatches: 17

Query Match: 91.39% Indels: 0

DB: 6 Gaps: 0

US-09-672-725C-6 (1-368) x AX108656 (1-4195)

Qy 1 ThrSerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArg 20

Db 1780 ACGTACGCCCTTGACACAGAAAGTGAAGCAGTGTTCAGGTGGCTTGGATAGGCCAGA 1839

Qy 21 LysGlyArgThrThrIleAlaIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40

Db 1840 AAAGGTGGACCCACCATGTGTATAGTATCATCGTTGTCTACGGTTCGTATGCCGAGCTC 1899

Qy 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60

Db 1900 ATCGCTGTTTCGATGATGAGTCATTTGTGGAGAAAGGAAATCATGATGAGCTCATGAAA 1959

Qy 61 GluLysGlyIleIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80

Db 1960 GAGAAAGGCATTTACTTCAAACTTGTCAATGCACAGCAGCAAGAAATGAAATTTGAATTA 2019

Qy 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuLeuMetSerProLysAsp 100

Db 2020 GAAATGCACCTGATGAATCCAAAAGTGAATTTGATACCTTGGAAATGCTTCCACATGAT 2079

Qy 101 SerGlySerSerLeuLysLysArgSerThrArgArgSerIleHisAlaProGlnGly 120

Db 2080 TCAGGATCCAGTCTAATAAGAAAAAGATCCACTCGTAGGAGTGTCCGTGGATCACAAGGC 2139

Qy 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140

Db 2140 CAAGACAGAAAGCTTAGTACCAGAGGCTCTGGATGAAAGTATACCTCCAGTTTCCTTT 2199

Qy 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160

Db 2200 TGGAGGATTAAGAAGCTTAATAATTAACACTGAGTGGCCTTATTTTGTGTGTGTTATTTGT 2259

Qy 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180

Db 2260 GCCATTATAAATGAGGCTCTGCAACACAGCATTTTGCAGTAATATATTTCAAAAGATTATAGG 2319

Qy 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200

Db 2320 ATTTTTCAGAAATGATGATGCGCAACAAACAGACAGCAATAGTAACTTGTGTTTCACTA 2379

Qy 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220

Db 2380 TTGTTTCTAGTCTTGGAAATGTTTCTTTTATTACATTTTTCCTTCAGGGCTTCACATTT 2439

Qy 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240

Db 2440 GGCAAGCTGGAGAGATCCTCACCAACGGCTCCGATACATGTTTTCGATCCATGCTC 2499

Qy 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260

Db 2500 AGACAGGATGTGAGCTGTTGATGACCCCTAAAACACCACTGGAGCATTTGACTACCAGG 2559

Qy 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280

Db 2560 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCCAGGCTTCGATATAATATACC 2619

Qy 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300

Db 2620 CAGAAATAGCAAAATCTTGGACAGGAATAATATATATCTTAATCTATGTTGGCAACTG 2679

Qy 301 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320

Db 2680 ACATGTTACTCTTAGCAATTTGTACCCATCATTCATAGCAGGAGTGTGTGAAATGAAA 2739

Qy 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340

Db	2740	ATGTTGCTCGGACAAAGCACTGAAAGATAAGAAAGAACTAGAAAGTGCTGGGAAGATCGCT	2799
QY	341	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu	360
Db	2800	ACTGAAGCAATAGAAACTTCGGAAGTGTGTTCTTGGACTCAGGACGAGAAGTTGAA	2859
QY	361	HisMetTyrAlaGlnSerLeuGln	368
Db	2860	CATATGTATGATCAGAGTTTCAG	2883
RESULT 15			
I33621			
LOCUS	I33621	2726 bp	DNA linear PAT 06-FEB-1997
DEFINITION	Sequence 1 from patent US 5593840.		
ACCESSION	I33621		
VERSION	I33621.1	GI:1824412	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 2726)		
TITLE	Bhatnagar, S. K., George, A. L., Jr. and Nazarenko, I.		
JOURNAL	Amplification of nucleic acid sequences		
FEATURES	Patent: US 5593840-A 1 14-JAN-1997;		
	Location/Qualifiers		
	1..2726		
	/organism="unknown"		
BASE COUNT	818 a 538 c 642 g 728 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	2..16e-142	Length:	2726
Score:	1672.00	Matches:	333
Percent Similarity:	94.84%	Conservative:	16
Best Local Similarity:	90.49%	Mismatches:	19
Query Match:	91.07%	Indels:	0
DB:	6	Gaps:	0
US-09-672-725C-6 (1-368) x I33621 (1-2726)			
QY	1	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	20
Db	176	ACGTACGCTTGGACACAGAAAGCGAAGCAGTGGTTCAGGTGGCTCTGGATAAGGCCAGA	235
QY	21	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	40
Db	236	AAAGTCGGACCACTGATGATAGCTCATCGTTGCTACAGTTCGTAAATGCTGACGTC	295
QY	41	IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys	60
Db	296	ATCGCTGGTTTCGATGATGGAGTCATTGTGGAGAAAGAAATCATGATGAACTCATGAA	355
QY	61	GluLysGlyIleTyrPheLysLeuValThrMetClnThrArgGlyAsnGluIleAspLeu	80
Db	356	GAGAAAGCAATTTACTTCAACTTGTCCAAATGCAGACAGCAGGAATGAAGTTGAATTA	415
QY	81	GluAsnAlaThrGlyGluSerLysSerGluSerAlaLeuGluMetSerProLysAsp	100
Db	416	GAATAATGCAGCTGATGAATCCAAAGTGAATTTGATGCCCTTGGAAATGCTTCAAAATGAT	475
QY	101	SerGlySerSerLeuLysArgSerThrArgSerIleHisAlaProGlnGly	120
Db	476	TCAAGATCCAGCTTAATAAGAAAGAACTCACTCGTAGGAGTGTCCGTGGATCACAAGCC	535
QY	121	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe	140
Db	536	CAGACAGAAGCTTAGTACCAAGAGGCTCTGGATGAAGTATACCTCCAGTTTCCTTT	595
QY	141	TrpArgIleLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	160
Db	596	TGGAGGATTATGAGCTAAATTTAACTGAATGGCCCTATTGTTGTTGGTATATTTGT	655
QY	161	AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	180

Db	656	GCCATTATAAATGGAGCCCTGCACACGCAATTTGCATAATAATTTTCAAAAGATTATAGG	715
QY	181	IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal	200
Db	716	GTTTTACAAGAAATTGATGATCTCTGAAACAAACACAGCAATAGTAACITGTTTTCAC	775
QY	201	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	220
Db	776	TTGTTCTAGCCCTTGGAAATTTCTTTTATTACATTTTCTTCCCTCAGGGTTTTCACAT	835
QY	221	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	240
Db	836	GGCAAGCTGGAGAGATCTCTCACCACGGGCTCCGATACATGTTTTCGATCCATGCTC	895
QY	241	ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrArg	260
Db	896	AGACAGGATGTGAGTTGGTTTGATACCCCTAAAAACACCACCTGGAGCAATTGACTAC	955
QY	261	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	280
Db	956	CTGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTCCAGGCTTCTGTAAATACC	1015
QY	281	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	300
Db	1016	CAGATATAGCAATCTTGGGACAGGAATAATTATATCTTTCATCTATGTTGGCAACTA	1075
QY	301	ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	320
Db	1076	ACACTGTTACTCTTAGCAATTTGACCATCATTTGCAATAGCAGGAGTTCTTGAATGAA	1135
QY	321	MetLeuSerGlyClnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla	340
Db	1136	ATGTTGCTCTGGACAAAGCACTGAAAGATAAGAAAGAACTAGAAAGGTCTGGGAAG	1195
QY	341	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu	360
Db	1196	ACTGAAGCAATAGAAAACCTCCGAAACCGTTGTTCTTTCAGTCAGGACGAGGTTT	1255
QY	361	HisMetTyrAlaGlnSerLeuGln	368
Db	1256	CATATGATGCTCAGAGTTTTCAG	1279

Search completed: November 7, 2002, 04:53:59
Job time : 1138.74 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:36:49 ; Search time 86.6553 seconds
(without alignments)
7291.243 Million cell updates/sec

Title: US-09-672-725C-6

Perfect score: 1836

Sequence: 1 TSALDTSEAVVQVLDKAR.....TVVSLTQEQKFEHMYAQSILQ 368

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09672725/runat_04112002_124338_17877/app_query.fasta_1.6435
-DB=N Geneseq_032802 -QPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THRESHOLD=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725 -CGN_1_1_0 @runat_04112002_124338_17877 -NCPD=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	1836	100.0	1107	22	AAD03490	Dog P-glycoprotein
2	1808	98.5	4279	22	AAD03488	Dog P-glycoprotein
3	1808	98.5	4279	22	AAD03504	Dog P-glycoprotein
4	1808	98.5	4279	22	AAD03505	Dog P-glycoprotein
5	1808	98.5	4279	22	AAD03506	Dog P-glycoprotein
6	1800	98.0	4317	22	AAD03489	Dog P-glycoprotein
7	1678	91.4	4186	22	AAF86127	Cynomolgous monke
8	1678	91.4	4195	22	AAF86128	Cynomolgous monke
9	1672	91.1	2726	15	AAQ70907	Multidrug-resistan
10	1672	91.1	2726	15	AAQ70916	Multidrug-resistan
11	1672	91.1	2726	18	AAQ73322	Multidrug resistan
12	1672	91.1	3860	21	AA249332	Human wild-type mu
13	1672	91.1	3860	21	AA249333	Human G185V mutant
14	1672	91.1	3860	24	ABA94365	Human BCRP DNA rel
15	1672	91.1	3860	24	ABA94366	Human BCRP DNA rel
16	1672	91.1	3988	21	AA288973	Human MDR-1 DNA.
17	1672	91.1	4264	19	AAV66533	Mutated human P-gl
18	1672	91.1	4264	19	AAV66534	Mutated human P-gl
19	1672	91.1	4349	22	AAH57442	Human intestine ce
20	1672	91.1	4646	15	AAQ72872	Human multidrug re
21	1672	91.1	4646	21	AA294738	Human ATP binding
22	1672	91.1	4669	8	AAH70752	Sequence of human
23	1672	91.1	6505	17	AAT13394	Hybrid vector pSF-
24	1672	91.1	8630	21	AA224041	Retroviral M4 mdr-
25	1672	91.1	8630	21	AA224042	Retroviral vector
26	1669	90.9	4378	11	AAQ04522	Multidrug Resistan
27	1665	90.7	4669	14	AAQ52726	Sequence of human
28	1644	89.5	4669	19	AAV32845	Human P glycoprote
29	1584	86.3	4788	21	AAZ49335	Murine multidrug r
30	1584	86.3	4788	24	ABA94368	Mouse BCRP DNA rel
31	1575	85.8	4369	21	AAZ52047	Rat multidrug resi
32	1575	85.8	4425	21	AAZ52048	Rat multidrug resi
33	1426.5	77.7	4233	21	AAZ90198	Rat mdr1b2 (multis
34	1426.5	77.7	4233	22	AAF27498	Murine multidrug r
35	1401.5	76.3	4189	21	AAZ49334	Mouse BCRP DNA rel
36	1401.5	76.3	4189	24	ABA94367	Mouse multidrug re
37	1401.5	76.3	4313	14	AAQ38950	Human ATP binding
38	1248.5	68.0	3924	21	AAZ94742	Human MDR-3 DNA.
39	1248.5	68.0	3924	21	AAZ88974	Human ATP binding
40	788.5	42.9	4776	21	AAZ94744	H. contortus PGP-A
41	677.5	36.9	2698	20	AAV69394	H. contortus PGP-A
42	677.5	36.9	4175	20	AAV69392	H. contortus PGP-B
43	667	36.3	2681	20	AAV69396	Drosophila melanog
44	660	35.9	3942	23	ABL14175	H. contortus PGP-O
45	655.5	35.7	3512	20	AAV69395	

ALIGNMENTS

RESULT 1
AAD03490
ID AAD03490 standard; cDNA; 1107 BP.
XX
AC AAD03490;
XX
XX 13-JUN-2001 (first entry)
DT
DT Dog P-glycoprotein (PGP) cDNA #3.
DE
DE Dog; P-glycoprotein; PGP; multidrug transporter; MDRI;
KW drug bioavailability; transgenic animal; genetic model; ss.
XX
XX Canis familiaris.
OS
XX
FH Key Location/Qualifiers
CDS 1..1104
FT /*tag= a

FT	/product="Dog P-glycoprotein (PGP). #3"
FT	/note="The coding region does not include stop codon"
FT	/partial
PN	
XX	WO200123540-A2.
XX	
PD	05-APR-2001.
XX	
PF	28-SEP-2000; 2000WO-US26767.
XX	
PR	28-SEP-1999; 99US-0156510.
XX	
PA	(GENT-) GENTEST CORP.
XX	
PI	Stocker PJ, Stedmel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX	
DR	WPI; 2001-235373/24.
DR	P-PSDB; AAE00305.
XX	
PT	New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
PT	for determining the bioavailability of drugs and for screening for dog
PT	PGP inhibitors -
XX	
PS	Claim 1; Page 75-77; 11pp; English.
XX	
CC	The invention relates to dog P-glycoprotein (PGP) also referred
CC	as multidrug transporter (MDR1) and nucleic acids encoding them.
CC	The invention also includes fragments and biologically functional
CC	variants of dog P-glycoprotein. PGP and their nucleic acids are
CC	useful for determining the bioavailability of drugs and for
CC	screening PGP inhibitors. They are useful for the diagnosis and
CC	treatment of conditions characterised by PGP activity, by
CC	reducing or increasing PGP activity in a cell. PGP nucleic acids
CC	are used as oligonucleotide probes. Complements of PGP nucleic
CC	acids are useful as antisense oligonucleotides, to induce a PGP
CC	'knockout' phenotype. They are used to prepare a non-human
CC	transgenic animal, which are valuable as genetic models for
CC	human diseases.
CC	The present sequence is dog P-glycoprotein (PGP) cDNA. The
CC	PGP enzyme functions as an efflux pump exporting small molecules
CC	across the cell membrane. This enzyme is a member of the ABC
CC	transporter family.
XX	
SQ	Sequence 1107 BP; 341 A; 200 C; 256 G; 310 T; 0 other;
Alignment Scores:	
Pred. No.:	9, 86e-211 Length: 1107
Score:	1836.00 Matches: 368
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	22 Gaps: 0
US-09-672-725C-6 (1-368) x AAD03490 (1-1107)	
QY	1 ThrSeraLaLeuAspThrGluSerGluAlaValValcInValAlaLeuAspLysAlaArg 20
Db	1 ACGTCAGCTCTGGACACTGAAAGTGAAGCAGTGTTCAGGTGGCCCTGGATAAGCCAGA 60
QY	21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db	61 AAGGCCCGACTACCATTTGGTATGAGTCATCGTTTGCTACAGTTCTGAATGCCGATGTC 120
QY	41 IleAlaGlyPheAspAspGlyValIleValcLulysGlyAsnHisAspGluLeuMetLys 60
Db	121 ATTGCTGGTTTGTATGGATGAGTCATTGTGAGAAAGAAATCATGATGAACATGAA 180
QY	61 GluLysGlyIleTyrPheLysLeuValThrMetcInThrArgGlyAsnGluIleAspLeu 80
Db	181 GAGAAGGCAATTTACTTCAAACCTTGTCACATGTCAGACAAAGAGAAATGAAATTTGACTTA 240
QY	81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100

D	b	241	GAAAAATGCCACTGGTGAATCCAAAAGTGAAAGTGATGCGCTTGGAAATGTCTCCAAGAAT	300
Q	y	101	SerGlySerLeuLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly	120
D	b	301	TCAGGGTCCAGTTTATAAANAAGAGATCAACTCGCAGGAGTAGTACATGCACCACAAGGC	360
Q	y	121	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	140
D	b	361	CAAGACAGAAAAGCTTGGTACAAAAGAGACTTGAATGAGAATGCTACCTCCAGTTTCCCTTC	420
Q	y	141	TripArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	160
D	b	421	TGGAGGATTCTGAAGCTGAACCTCACTGAATGGCCCTATTATTGGTGGTATATTTTGT	480
Q	y	161	AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	180
D	b	481	GCTATTATAAACGAGCGCTGCAGCCAGCAATTTTCAATAATATATTTTCAAGGATTATAGG	540
Q	y	181	IlePheThrArgaspGluaspProGlnThrLysArgGlnIleSerAsnMetPheSerVal	200
D	b	541	ATCTTTTACCAGCATGAGGATCCTCGAAACAACACACAGANTAGTAACATGTTTTCTGTA	600
Q	y	201	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	220
D	b	601	TTGTTTCTAGCTCTGGAATATTCTTTTATTACATTTTTCCTTCAGGGTTTCACATTT	660
Q	y	221	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	240
D	b	661	GGCAAAGCTGGAGAGATCCTCACTAAGCGGCTTCGATACATGGTTTTTCAGATCCATGCTG	720
Q	y	241	ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg	260
D	b	721	ACACAGCATGTCAGCTGGTTGATGACCTTAAAAACACCCTGGAGCATTTGACCACCAGG	780
Q	y	261	LeuAlaasnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	280
D	b	781	CTTGCCAAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGCTTGCTGTCAITACC	840
Q	y	281	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleIleSerLeuIleTyrGlyTrpGlnLeu	300
D	b	841	CAGAAATAGCAAACTCTGGCAGAGCATTATTATATCTTAATCTATGTTGGCAATTA	900
Q	y	301	ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	320
D	b	901	ACACTTTTACTCTTAGCAATTTGACCCATCATTGCATAGCAGGAGTTGTTGAARTGAAA	960
Q	y	321	MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla	340
D	b	961	ATGTTGTCGGACCAAGCACCTGAAAGATAAGAAAGAGCTAGAAGGAGCTGGGAAGATTGCT	1020
Q	y	341	ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu	360
D	b	1021	ACAGAAGGCATPCGAAAACCTTCCGAAGCTGTGTTTCTTGTACTCAGGAGCAGAAGTTTGA	1080
Q	y	361	HisMetTyrAlaGlnSerLeuGln	368
D	b	1081	CACATGTATGCACAGATTTGCAG	1104
RESULT 2				
ID	AAD03488			
XX	AAD03488 standard; cdna; 4279 BP.			
AC	AAD03488;			
XX				
DT	13-JUN-2001 (first entry)			
XX				
DE	Dog P-glycoprotein (PGP) cDNA #1.			
XX				
KW	Dog; p-glycoprotein; PCP; multidrug transporter; MDR1;			
OS	drug bioavailability; transgenic animal; genetic model; ss.			
XX				
OS	Canis familiaris.			
XX				

Key Location/Qualifiers
 CDS 17..3862
 FT /*tag= a
 FT /product= "Dog P-glycoprotein (PGP) #1"
 XX

PN WO200123540-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26767.

XX 28-SEP-1999; 99US-0156510.

XX (GENT-) GENTEST CORP.

XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 WPI; 2001-235373/24.

XX P-PSDB; AAE00303.

PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 for determining the bioavailability of drugs and for screening for dog
 PGP inhibitors -

PS Claim 3; Page 58-63; 11lpp; English.

CC The invention relates to dog P-glycoprotein (PGP) also referred
 as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 variants of dog P-glycoprotein. PGP and their nucleic acids are
 useful for determining the bioavailability of drugs and for
 screening PGP inhibitors. They are useful for the diagnosis and
 treatment of conditions characterised by PGP activity, by
 reducing or increasing PGP activity in a cell. PGP nucleic acids
 are used as oligonucleotide probes. Complements of PGP nucleic
 acids are useful as antisense oligonucleotides, to induce a PGP
 'knockout' phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 human diseases.

CC The present sequence is dog P-glycoprotein (PGP) cDNA. This
 sequence is also referred as Genotype C cDNA. The
 CC PGP enzyme functions as an efflux pump exporting small molecules
 across the cell membrane. This enzyme is a member of the ABC
 CC transporter family.

XX SQ Sequence 4279 BP; 1294 A; 834 C; 1008 G; 1143 T; 0 other;

Alignment Scores:

Pred. No.:	1.87e-206	Length:	4279
Score:	1808.00	Matches:	363
Percent Similarity:	99.46%	Conservative:	3
Best Local Similarity:	98.64%	Mismatches:	2
Query Match:	98.47%	Indels:	0
DB:	22	Gaps:	0

US-09-672-725C-6 (1-368) x AAD03488 (1-4279)

QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValAlnValAlaLeuAspLysAlaArg 20
 DB 1694 ACCTCAGCTCGACACTGAAAGTGAGCGTGGTTTCAGTGGCCCTGGATAAGGCCAGA 1753
 QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
 DB 1754 AAAGCGCGGACTACCATTTGTATAGCTCATCGTTTGTCTACAGTTCGTAATGCCGATGC 1813
 QY 41 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
 DB 1814 ATTGCTGGTTTGTATGATGGAGTCATTGTGGAGAAAGGAATCATGATGATCATGAAA 1873
 QY 61 GluLysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
 DB 1874 GAGAAGGGCATTTACTTCAAACTTGTCAATGTCACAGACAGAGGAATGAAATTTGAGTTA 1933

QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
 DB 1934 GAAATGCCACTGGTGAATCCAAAAGTGAAGTATGCTTGGAAATGTCTCCAAAAGAT 1993
 QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
 DB 1994 TCAGGTCCTCAGTTTAATAAAAAAGAGATCAACTCGCAGGAGTATACATGCACACAGGC 2053
 QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
 DB 2054 CAAGACAGAAAGCTTGGTACAAAAGAGACTTGAATGAGAATGTACTCCAGTTCTCTTC 2113
 QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProThrPheValValGlyIlePheCys 160
 DB 2114 TGGAGGATTCGAAGCTCAACTCAACTGAATGGCCTTATTTTGTGGTGTATATTTGT 2173
 QY 161 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
 DB 2174 GCTATTATAACGAGCGCTGCAACAGCATTTTCAATAATATTTCAAGGATTATAGGG 2233
 QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
 DB 2234 ATCTTTACCCGAGATGAGGATCCTGAAACAAACACGACAGAAATAGTAACATGTTTCTGTA 2293
 QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
 DB 2294 TTGTTTCTAGTCCTTGGAAATATTTCTTTTATTACATTTTCTCTCAGGCTTCACATTT 2353
 QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
 DB 2354 GGCAAACTGGGAGATCCCTCACTAAGCGCTTCGATACATGGTTTTCAGATCCCATGCTG 2413
 QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
 DB 2414 AGACAGGATGTACGTGGTTGATGACCTTAAACACACCTGGAGCATTTGACACACAGG 2473
 QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
 DB 2474 CTGCCAATGATCGGCTCAAGTTAAAGGGCTTATAGTTCCAGGCTTCTCTCATTTACC 2533
 QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleThrGlyTrpGlnLeu 300
 DB 2534 CAGAAATAGCAAAATCTGGGACAGCATTTATATATCTTAATCTATGTTGGCAATTA 2593
 QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
 DB 2594 ACATTTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGAGATTGTTGAAATGAAA 2653
 QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 340
 DB 2654 ATGTTGCTCGACACAGCATGAAAGATAAGAAAGAGCTAGAAGGAGCTGGGAAGATTGCT 2713
 QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGlnLysPheGlu 360
 DB 2714 ACAGAAGCCATCGAAACTTCCGAACTGTTGTTCTTTGACTCGGAGCAGAGATTGCAA 2773
 QY 361 HisMetTyrAlaGlnSerLeuGln 368
 DB 2774 TACATGATGCACAGAGTTTGCAA 2797
 RESULT 3
 ID AAD03504
 XX AAD03504 standard; cDNA; 4279 BP.
 XX AAD03504;
 XX 13-JUN-2001 (first entry)
 XX Dog P-glycoprotein (PGP) allelic variant (Genotype A) cDNA.
 XX Dog; P-glycoprotein allelic variant; multidrug transporter; MDR1;
 KW drug bioavailability; transgenic animal; genetic model; ss.
 XX

```

OS Canis familiaris.
XX FH Key Location/Qualifiers
XX CDS 17..3862
FT FT /*tag= a
FT FT /product= "dog P-glycoprotein (PGP) allelic variant
FT FT (Genotype A) protein"
FT FT replace (607, C)
FT FT /*tag= b
XX PN WO200123540-A2.
XX XX
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26767.
XX PR 28-SEP-1999; 99US-0156510.
XX PA (GENT-) GENTEST CORP.
XX PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX WPI: 2001-235373/24.
XX DR P-PSDB; AAE00308.
XX XX
XX PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX PT for determining the bioavailability of drugs and for screening for dog
XX PT PGP inhibitors -
XX PS Claim 9; Page 85-90; 111pp; English.
XX CC The invention relates to dog P-glycoprotein (PGP) also referred
XX CC as multidrug transporter (MDR1) and nucleic acids encoding them.
XX CC The invention also includes fragments and biologically functional
XX CC variants of dog P-glycoprotein. PGP and their nucleic acids are
XX CC useful for determining the bioavailability of drugs and for
XX CC screening PGP inhibitors. They are useful for the diagnosis and
XX CC treatment of conditions characterised by PGP activity, by
XX CC reducing or increasing PGP activity in a cell. PGP nucleic acids
XX CC are used as oligonucleotide probes. Complements of PGP nucleic
XX CC acids are useful as antisense oligonucleotides, to induce a PGP
XX CC 'knockout' phenotype. They are used to prepare a non-human
XX CC transgenic animal, which are valuable as genetic models for
XX CC human diseases.
XX CC The present sequence is dog P-glycoprotein (PGP) allelic variant
XX CC (Genotype A) cDNA. The PGP enzyme functions as an efflux pump
XX CC exporting small molecules across the cell membrane. This enzyme
XX CC is a member of the ABC transporter family.
XX SQ Sequence 4279 BP; 1295 A; 833 C; 1008 G; 1143 T; 0 other;

Alignment Scores:
Pred. No.: 1.87e-206 Length: 4279
Score: 1808.00 Matches: 363
Percent Similarity: 99.46% Conservatives: 3
Best Local Similarity: 98.64% Mismatches: 2
Query Match: 98.47% Indels: 0
DB: 22 Gaps: 0

US-09-672-725C-6 (1-368) x AAD03504 (1-4279)
QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspPlysAlaArg 20
DB 1694 ACCTGAGCTCTGGACACTGAAAGTGAAGCAGTGGTTCAGGTGGCCCTGGATAAGGCCAGA 1753
QY 21 LysGlyArgThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
DB 1754 AAAAGCCGGACTACCATTTGTGTATAGCTCATCGTTTGTCTACAGTTCGTAATGCCGATGTC 1813
QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
DB 1814 ATTGCTGTTTGTGATGATGAGTCAATTTGGGAGAAGGAATCATGATGAACTCATGAA 1873

```

```

QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
DB 1874 GAGAAAGGCGCATTTACTTCAAACTTGTCAAAATGCAGACAGAGAAATGAAATGAGTTA 1933
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
DB 1934 GAAATGCCACTGGTGAATCCAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1993
QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
DB 1994 TCAGGCTCCAGTTTAATAAAGAAAGAGTCAACTCGCAGGAGTATACATGCACCAAGGC 2053
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
DB 2054 CAAGACAGAAAGCTTGGTACAAAGAGGACTTGAATGAAGTGAAGTGAAGTGAAGTGAAGT 2113
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValIlePheCys 160
DB 2114 TGGAGGATTCGAAGCTGAACCTCAACTCACTCACTCACTCACTCACTCACTCACTCACTCA 2173
QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
DB 2174 GCTATTATAACGGAGGCTGCAACCGAGCATTTTCAATAATATTTTCAAGGATTATAGGG 2233
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
DB 2234 ATCTTTACCGAGATGAGGATCTCTGAAACAAACACACAGATAGTAACTGTTTCTGTA 2293
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlnPheThrPhe 220
DB 2294 TTGTTTCTAGTCTCTTGGAAATTTCTTTTATTACATTTTCTCCAGGCTTCACATTT 2353
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
DB 2354 GGCAAAGCTGGGAGATCTCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTG 2413
QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
DB 2414 AGACAGGATGTGACGTGGTGTGATGACCCCTAAACACACCTGGAGCATTGACACACAGG 2473
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleIleGlySerArgLeuAlaValIleThr 280
DB 2474 CTGGCAATGATCGGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGTCTCATTTACC 2533
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
DB 2534 CAGATATATAGCAAACTCTGGGACAGGCATTTATATATCTTAATCTATGTTGGCAATTA 2593
QY 301 ThrLeuLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
DB 2594 ACATTTTCTCTTAGCAATTTGACCCATCATTTGCAATAGCAGGAGTTCTTGAATGAAA 2653
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
DB 2654 ATCTTGTCTGGCAAGCACTGAAAGATAGAAAGAGAGCTAGAAAGAGCTGGGAGAGTTGCT 2713
QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
DB 2714 ACAGAAAGCCATCGAAACCTCCGAACCTGTTTCTTCTGCTCGGAGCAGAAAGTTTGA 2773
QY 361 HisMetTyrAlaGlnSerLeuGln 368
DB 2774 TACATGATGCACAGAGTTTGCAA 2797
RESULT 4
AAD03505
ID AAD03505 standard; cDNA; 4279 BP.
XX AC AAD03505;
XX XX 13-JUN-2001 (first entry)
XX XX Dog P-glycoprotein (PGP) allelic variant (Genotype B) cDNA.

```

```

XX Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
KW MDRI; drug bioavailability; transgenic animal; genetic model; ss.
XX
OS Canis familiaris.
XX
FH Key Location/Qualifiers
FT CDS 17..3862
FT     /*tag= a
FT     /product= "Dog P-glycoprotein (PGP) allelic variant
FT     (Genotype B) protein"
FT allele replace (91, T)
FT     /*tag= b
FT allele replace (607, C)
FT     /*tag= c
XX
XX WO200123540-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26767.
XX
XX 28-SEP-1999; 99US-0156510.
XX
XX (GENT-) GENTEST CORP.
XX
XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX WPI; 2001-235373/24.
XX P-PSDB; AAE00309.
XX
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX for determining the bioavailability of drugs and for screening for dog
XX PGP inhibitors -
XX
XX Claim 9; Page 93-99; 111pp; English.
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
XX as multidrug transporter (MDRI) and nucleic acids encoding them.
XX The invention also includes fragments and biologically functional
XX variants of dog P-glycoprotein, PGP and their nucleic acids are
XX useful for determining the bioavailability of drugs and for
XX screening PGP inhibitors. They are useful for the diagnosis and
XX treatment of conditions characterised by PGP activity, by
XX reducing or increasing PGP activity in a cell. PGP nucleic acids
XX are used as oligonucleotide probes. Complements of PGP nucleic
XX acids are useful as antisense oligonucleotides, to induce a PGP
XX 'knockout' phenotype. They are used to prepare a non-human
XX transgenic animal, which are valuable as genetic models for
XX human diseases.
XX
XX The present sequence is dog P-glycoprotein (PGP) allelic variant
XX (genotype B) cDNA. The PGP enzyme functions as an efflux pump
XX exporting small molecules across the cell membrane. This enzyme
XX is a member of the ABC transporter family.
XX
XX Sequence 4279 BP; 1296 A; 833 C; 1008 G; 1142 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1..87e-206 Length: 4279
XX Score: 1808.00 Matches: 363
XX Percent Similarity: 99.46% Conservative: 3
XX Best Local Similarity: 98.64% Mismatches: 2
XX Query Match: 98.47% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-672-725c-6 (1-368) x AAD03505 (1-4279)
XX
XX QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
XX |||||||
XX Db 1694 ACGTGAGCTGGACACTGAAAGTGAAGCAGTGGTTTCAGGTGGCCCTGGATAAGGCCAGA 1753
XX |||||||
XX QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValValArgAsnAlaAspVal 40
XX |||||||

```

```

Db 1754 AAAGCGCGGACTACCATTTGTGTAGCTCATCGTTTCTGTACAGTTCTGTAATGCCGATGTC 1813
QY 41 IleAlaGlyPheAspAspGlyValIleValGluGlyGlyAsnHisAspGluLeuMetLys 60
|||
Db 1814 ATTGCTGGTTTGTATGGAGTCAATTGTGGAGAAAGAAATCATGATGAATCATGAATA 1873
QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
|||
Db 1874 GAGAGGGCAATTTACTTCAACATTGTCACATGACAGAGAGAAATGAAATTTAGTTA 1933
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
|||
Db 1934 GAAATGCCACTGGTGAATCCAAAGTGAAGTGCCTTGGAAATGCTCCAAAGAT 1993
QY 101 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 120
|||
Db 1994 TCAGGGTCCAGTTTAAATAAAGAGATCAACTCGCAGGAGTATACATCACCAAGGC 2053
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
|||
Db 2054 CAAGACAGAAAGCTTGGTACAAAGAGAGACTTGAATGAGAATGTACCTCCAGTTCTCTTC 2113
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
|||
Db 2114 TGGAGGATTTCTGAAGCTGAACCTCAACTGAATGGCCTTATTTTGGTGGTATATTTGT 2173
QY 161 AlaIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
|||
Db 2174 GCTATTATAACGGAGGCGCTGCAACACGACATTTTCAATAATATTTTCAAGGATTATAGG 2233
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
|||
Db 2234 ATCTTTACCCGAGATGAGGATCCTGAAACAAACACACAGATAGTAACATGTTTCTGTA 2293
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
|||
Db 2294 TTGTTTCTAGTCTTGGAAATTTATTTTATTACATTTTCTCTCAGGGCTTCACATTT 2353
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
|||
Db 2354 GGCAAAGCTGGGGAGATCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTG 2413
QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
|||
Db 2414 AGACAGGATGTCAGCTGGTTTCATGCCCTAAACACACCCTGGAGCATTGACACACAGG 2473
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
|||
Db 2474 CTTGCCAATGATGCGGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGTGTCATTACC 2533
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
|||
Db 2534 CAGAATATAGCAAAATCTTGGGACAGGCATTTATATCTTAATCTATGGTTGGCAATTA 2593
QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValValGluMetLys 320
|||
Db 2594 ACACATTTTACTCTTAGCAATTTGACCCATCATTCGCAATAGCAGGAGTTGTTGAATGAA 2653
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGlyAlaGlyLysIleAla 340
|||
Db 2654 ATGTTGCTGGACAAAGCAGCTGAAAGATAGAAAGAGCTAGAAAGAGAGCTGGGAAGATTGCT 2713
QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlyPheGlu 360
|||
Db 2714 ACAGAAAGCCATCGAAACTTCCGAACTGTTGTTCTTTGACTCGGGAGCAGAAAGTTTGA 2773
QY 361 HisMetTyrAlaGlnSerLeuGln 368
|||
Db 2774 TACATGTATGCACAGAGTTTGCAA 2797
XX
XX RESULT 5
XX AAD03506
XX ID AAD03506 standard; cDNA; 4279 BP.

```

XX AAD03506;
AC
XX 13-JUN-2001 (first entry)
XX
XX Dog P-glycoprotein (PGP) allelic variant (Genotype D) cDNA.
XX
XX Dog; P-glycoprotein allelic variant; PGP; multidrug transporter;
KW MDRI; drug bioavailability; transgenic animal; genetic model; ss.
XX
XX Canis familiaris.
XX
XX Key Location/Qualifiers
FT CDS 17..3862
FT /*tag= a
FT /product= "dog P-glycoprotein (PGP) allelic variant
FT (Genotype D) protein"
FT allele replace (91, T)
FT /*tag= b
FT allele replace (607, C)
FT /*tag= c
FT allele replace (1001, T)
FT /*tag= c
FT allele replace (3458, A)
FT /*tag= c
XX W0200123540-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26767.
XX
XX 28-SEP-1999; 99US-0156510.
XX
XX (GENT-) GENTEST CORP.
XX
XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX
XX WPI; 2001-235373/24.
XX P-PSDB; AAE00310.
XX
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX for determining the bioavailability of drugs and for screening for dog
XX PGP inhibitors -
XX
XX Claim 9; Page 102-107; 11lpp; English.
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
XX as multidrug transporter (MDRI) and nucleic acids encoding them.
XX The invention also includes fragments and biologically functional
XX variants of dog P-glycoprotein. PGP and their nucleic acids are
XX useful for determining the bioavailability of drugs and for
XX screening PGP inhibitors. They are useful for the diagnosis and
XX treatment of conditions characterised by PGP activity, by
XX reducing or increasing PGP activity in a cell. PGP nucleic acids
XX are used as oligonucleotide probes. Complements of PGP nucleic
XX acids are useful as antisense oligonucleotides, to induce a PGP
XX 'knockout' phenotype. They are used to prepare a non-human
XX transgenic animal, which are valuable as genetic models for
XX human diseases.
XX The present sequence is dog P-glycoprotein (PGP) allelic variant
XX (Genotype D) cDNA. The PGP enzyme functions as an efflux pump
XX exporting small molecules across the cell membrane. This enzyme
XX is a member of the ABC transporter family.
XX
XX Sequence 4279 BP; 1296 A; 833 C; 1009 G; 1141 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 1.87e-206 Length: 4279
Score: 1808.00 Matches: 363
Percent Similarity: 99.46% Conservative: 3
Best Local Similarity: 98.64% Mismatches: 2
Query Match: 98.47% Indels: 0

DB: 22 Gaps: 0
US-09-672-725C-6 (1-368) x AAD03506 (1-4279)
Qy 1 ThrSerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArg 20
Db 1694 ACGTCAGCTCTGGACACTGAAAGTGAAGCAGTGGTTTCAGGTGGCCCTGGGTAAGCCGAGA 1753
Qy 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1754 AAAGCCGACTACATCTGTAGTACCTCATCGTTTGTCTACAGTCTGTAATGCCGATGTC 1813
Qy 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1814 ATTGCTGTTTGTATGATGGAGTCAATGCTGAGAAAGAAATCATGATGAACATCATGAAA 1873
Qy 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
Db 1874 GAGAAGGCACTTTACTTCAAACTTGTCAACATGCAGACAAGAGAAATGAAATTTGAGTTA 1933
Qy 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 1934 GAAATGCCACTGGTGAATCCAAAAGTGAAGTGCCTTGGAAATGCTCTCCAAAAGAT 1993
Qy 101 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 120
Db 1994 TCAGGTCCTAGTTTAAATAAAAGAAAGATCAACTCGCAGGAGTATACATGCCACCAAGGC 2053
Qy 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 140
Db 2054 CAAGACAGAAAGCTTGGTACAAAGAGGACTTGAATGAGAATGTACCTCCAGTTTCCTTC 2113
Qy 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
Db 2114 TGGAGGATTCGAAGCTCAACTGAATGGCTTATTTTGTGGTGTATATTGT 2173
Qy 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleGly 180
Db 2174 GCTATTATAAACGGAGGCTGCAACACAGCATTTTCAATAATATTTTCAAGGATTATAGG 2233
Qy 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2234 ATCTTTACCCGAGATGAGATCCTGAAACAAACGACAGATAGTAACATCTTTCTGTA 2293
Qy 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPheLeuGlnGlyPheThrPhe 220
Db 2294 TTGTTTCTAGTCTCTTGAATATTTCTTTTATTTACATTTTCTCTCCAGGGCTTCACATT 2353
Qy 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2354 GGCAAGCTGGGGAGATCCTCACTAAGCGCTTCGATACATGGTTTCAGATCCATGCTG 2413
Qy 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 2414 AGACAGGATGTCAGCTGGTTGATGACCTTAAACACACCTGGAGCATTTGACAACACAGG 2473
Qy 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2474 CTTGCCAATGATGCGGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGCCTGCTATTACC 2533
Qy 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 2534 CAGAATATAGCAAAATCTTGGGACAGCATATATATATCTTAATCTATGGTTGCCAATA 2593
Qy 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
Db 2594 ACACITTTACTCTTAGCAATTTGCCATCATTCACCATTCATTCGCAATAGCAGGAGTTGTTGAAATGAAA 2653
Qy 321 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 2654 ATGTTTGTCTGGACACGACTGAAAGATAGAAAGAGCTAGAAAGAGCTGGGAGGATTGCT 2713
Qy 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360

```
|||||
Db 2714 ACAGAAGCCATCGAAATTCGAACTGTGTTCTTGTGACTCGGAGCAGAGTTTGA 2773
QY 361 HisMetTyrAlaGlnSerLeuGln 368
Db 2774 TACATGTATGCACAGAGTTTGA 2797
RESULT 6
ID AAD03489
XX AAD03489 standard; cDNA; 4317 BP.
AC AAD03489;
XX
DT 13-JUN-2001 (first entry)
DE Dog P-glycoprotein (PGP) cDNA #2.
XX
KW Dog; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW drug bioavailability; transgenic animal; genetic model; ss.
XX
OS Canis familiaris.
XX
FH Key Location/Qualifiers
FT CDS 70..3912
FT /*tag= a
FT /product= "Dog P-glycoprotein (PGP) #2"
PN WO200123540-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26767.
XX
PR 28-SEP-1999; 99US-0156510.
XX
PA (GENT-) GENTEST CORP.
PI
XX Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
XX WPI; 2001-235373/24.
DR P-PSDB; AAE00304.
XX
XX New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
XX for determining the bioavailability of drugs and for screening for dog
XX PGP inhibitors
XX
XX Claim 1; Page 66-72; 111pp; English.
XX
XX The invention relates to dog P-glycoprotein (PGP) also referred
XX as multidrug transporter (MDR1) and nucleic acids encoding them.
XX The invention also includes fragments and biologically functional
XX variants of dog P-glycoprotein. PGP and their nucleic acids are
XX useful for determining the bioavailability of drugs and for
XX screening PGP inhibitors. They are useful for the diagnosis and
XX treatment of conditions characterised by PGP activity, by
XX reducing or increasing PGP activity in a cell. PGP nucleic acids
XX are used as oligonucleotide probes. Complements of PGP nucleic
XX acids are useful as antisense oligonucleotides, to induce a PGP
XX 'knockout' phenotype. They are used to prepare a non-human
XX transgenic animal, which are valuable as genetic models for
XX human diseases.
XX
XX The present sequence is dog P-glycoprotein (PGP) cDNA. The
XX PGP enzyme functions as an efflux pump exporting small molecules
XX across the cell membrane. This enzyme is a member of the ABC
XX transporter family.
XX
XX Sequence 4317 BP; 1293 A; 844 C; 1019 G; 1161 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1-74e-205 Length: 4317
Score: 1800.00 Matches: 362
Percent Similarity: 99.18% Conservative: 3
Best Local Similarity: 98.37% Mismatches: 3
```

```
Query Match: 98.04% Indels: 0
DB: 22 Gaps: 0
US-09-672-725c-6 (1-368) x AAD03489 (1-4317)
QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
Db 1744 ACGTACAGTCTGGACACTGAAGTGAAGTGGTTCAGGTGGCCCTGGATAGGCCAGA 1803
QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1804 AAAGCCGGACTACCATTTGTATAGCTCATCGTTGTCTACAGTCTGTATGCCGATGC 1863
QY 41 IleAlaGlyPheAspAspGlyValIleValIleGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1864 ATTGCTGTTTGTATGATGAGTCAATTTGTGAGAAAGAAATCATGATGAACATCATGAA 1923
QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
Db 1924 GAGAGGGCATTTACTTCAAACTTGTCAATGCACAAGAGAAATGAATTGAGTTA 1983
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 1984 GAAATGCCACTGGTGAATCCAAAAGTGAAGTGATGCCCTTGAATGTCTCCAAAGAT 2043
QY 101 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 120
Db 2044 TCAGGGTCCAGTTTAAATAAAGAGATCAACTCGCAGGAGTATACATGCACCACAGGC 2103
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 140
Db 2104 CAAGACAGAAAGCTTGGTACAAAGAGGACTTGAATGAGAATGTTCTTCAGTTTCTTTC 2163
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
Db 2164 TGGAGGATTCGAAGCTGAACCTCAACTGAATGGCTTATTTTGTGGTGTATATTTCT 2223
QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 2224 GCTATTATAACGGAGGCTGCAACCAAGCATTTTCAATAATATTTTCAAGGATTTATAGG 2283
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2284 ATCTTTACCGAGATGAGGATCCTGAACAAACAGACAGCAATAGTAACATGTTTCTGTA 2343
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 2344 TTGTTTCTAGTCTCTGGAATTAATTTCTTTTATTATACATTTTCTCCAGGGCTTCACATTT 2403
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2404 GGCAAGCTGGGGAGATCCTCACTAAGCGGCTTCGATACATGTTTTCAGATCCATGCTG 2463
QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 2464 AGACAGAGATGTCAGCTGGTTTGTATGACCTTAAACACCACTGGAGCATTTGACACACAGG 2523
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2524 CTTGCCAATGATCGGGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGTCTGCTATTACC 2583
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 2584 CAGAAATATAGCAAACTCTGGGACAGGCATTAATATATCTTAATCTATGTTGGCAATTA 2643
QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
Db 2644 ACACITTTACTCTTAGCAATTTACCCATCATTTGCAATAGCAGAGATTGTTGAAATGAA 2703
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 2704 ATGTTGCTCTGGACAGCACTGAAGATAAGAAAGAGCTAGAAAGGAGCTGGGAAGATTGCT 2763
```

QY 341 ThrGluAlaIleGluAsnPhaArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
 DB 2764 ACAGAAGCCATCGAAACTCCGAACGTGTGTTCTTTCAGTCGGAGCAGAAGTTTGA 2823
 QY 361 HisMetTyrAlaGlnSerLeuGln 368
 DB 2824 TACATGTATGCAGAGTTTGCAA 2847
 RESULT 7
 AAF86127
 ID AAF86127 standard; cDNA; 4186 BP.
 XX AAF86127;
 AC
 XX
 XX
 DT 25-JUN-2001 (first entry)
 DE Cynomologous monkey P-glycoprotein cDNA.
 XX
 XX Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL;
 KW eflux pump; ss.
 KW
 XX
 OS Macaca fascicularis.
 XX
 FH Key Location/Qualifiers
 FT CDS 100..3942
 FT /*tag= a
 FT /product= "PGP"
 FT /note= "P-glycoprotein"
 PN W0200123565-A1.
 XX
 XX
 PD 05-APR-2001.
 XX
 XX 28-SEP-2000; 2000WO-US26592.
 XX
 XX 28-SEP-1999; 99US-0156921.
 PR 12-OCT-1999; 99US-0158818.
 XX
 XX (GENT-) GENTEST CORP.
 PA
 XX
 XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
 XI WPI; 2001-316136/33.
 DR P-PSDB; AAB81064.
 DR
 XX
 XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
 PT (PGP) and homologous PGP polypeptides are useful for predicting
 PT bioavailability of compound and increasing PGP transporter activity in
 cell
 XX
 XX Example 1; Page 51-57; 84pp; English.
 XX
 XX This invention relates to a polynucleotide sequence encoding a
 CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
 CC PGP protein. PGP, also known as multidrug transporter, MDRL is a member
 CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
 CC exporting small molecules across the cell membrane. The invention
 CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
 CC and protein, and also that of an allelic variant. The PGP polynucleotide
 CC sequence is useful for increasing PGP transporter activity in a cell.
 CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
 CC activity in a mammalian cell. They may also be used for increasing the
 CC bioavailability of a drug. The present sequence represents cDNA encoding
 CC cynomologous monkey P-glycoprotein.
 XX
 XX Sequence 4186 BP; 1226 A; 799 C; 1039 G; 1122 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 7.87e-191 Length: 4186
 Score: 1678.00 Matches: 333
 Percent Similarity: 95.38% Conservative: 18
 Best Local Similarity: 90.49% Mismatches: 17
 Query Match.: 91.39% Indels: 0

DB: 22 Gaps: 0
 US-09-672-725c-6 (1-368) x AAF86127 (1-4186)
 QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
 DB 1771 ACGTCAGCCTTGGACACAGAAAGTGAAGCAGTGGTTCAGGTGGCTCTGGATAAGCCAGA 1830
 QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
 DB 1831 AAGGTGGACACCACTGTGATAGCTCATCTGTTGTCTAGGGTTCGTAATGCCGCGTC 1890
 QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
 DB 1891 ATCGCTGCTTTCGATGATGGAGTCATCTGGAGAAAGGAAATCATGTAGCTCATGAAA 1950
 QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
 DB 1951 GAGAAAGCACTTACTTCAAACTTGTCAATGCAGACAGCAGGAATGAATTTGAATTA 2010
 QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
 DB 2011 GAAATGCAGCTGATGATCCAAAGTGAATTCATCTTGGAAATCTCTCATCATGAT 2070
 QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
 DB 2071 TCAGATCCAGTCTAATAAGAAAAAGATCCATCGTAGGAGTGGCCGTGGATCAACAGGC 2130
 QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
 DB 2131 CAAGACAGAAAGCTTAGTACCNAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 2190
 QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
 DB 2191 TGGAGGATATGAAGCTAAATTAACCTGAGTGGCCCTTATTTTGTGTGTATTTGT 2250
 QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
 DB 2251 GCCATTATAAATGGAGGCTCGCAACCCAGCATTTGCAACAGCATATATTTCAAGAGTATAGGG 2310
 QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
 DB 2311 ATTTTACAGAAGATGATGATGCCGAAACAAACACAGAAATAGTAACTTGTTCACATA 2370
 QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
 DB 2371 TTGTTTCTAGTCTTGGAAATGTTCTTTTATATACATTTTCTTCAGGGCTTCACATTT 2430
 QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
 DB 2431 GGCAAAGCTGGAGAGATCCTCACCAGGCGCTCCGATACATGGTTTCCGATCCATGCTC 2490
 QY 241 ArgGlnAspValSerTrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
 DB 2491 AGACAGGATGTGAGCTGGTTTGTATGACCCCTAAAAACACACCTGGAGCATTTGACTACCAG 2550
 QY 261 LeuAlaAsnAspAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
 DB 2551 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCAGGCTGCTATATATACC 2610
 QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
 DB 2611 CAGAAATAGCAAACTTGGGACAGGAATAATATATATCTTAATCTATGTGTGGCAACTG 2670
 QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
 DB 2671 ACACCTGTTACTCTTAGCAATTTACCCATTCATTCGAATAGCAGGAGTGTGTTGAATGAAA 2730
 QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
 DB 2731 ATGTTGTCGGACCAACACTGAAAGATAGAAAGAACTAGAGGTCTGGGAAGATCGCT 2790
 QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360

|||||
Db 2791 ACTGAAGCAATAGAAATTCGAACTGTGTTCTTGGACTCAGACAGAAAGTTTGA 2850
QY 361 HisMetTyrAlaGlnSerLeuGln 368
Db 2851 CATATGTATGATCAGAGTTTGAC 2874
RESULT 8
AAF86128
ID AAF86128 standard; cDNA; 4195 BP.
XX AAF86128;
XX 25-JUN-2001 (first entry)
XX Cynomologous monkey P-glycoprotein cDNA variant 1.
XX Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL1;
KW efflux pump; ss.
XX Macaca fascicularis.
OS
PH Key Location/Qualifiers
FT 100..3951
CDS /*tag= a
FT /product= "pgp"
FT /note= "P-glycoprotein"
FT misc_feature 376..384
FT /*tag= b
FT /note= "Insertion of 9 nucleotides relative to PGP
FT allelic variant AAF86127"
WO200123565-A1.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26592.
XX
XX 28-SEP-1999; 99US-0156921.
XX 12-OCT-1999; 99US-0158818.
XX
XX (GENT-) GENTEST CORP.
XX
XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
XX WPI; 2001-316136/33.
XX P-PSDB; AAB81065.
XX
XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
XX (PGP) and homologous PGP polypeptides are useful for predicting
XX bioavailability of compound and increasing PGP transporter activity in
XX cell -
XX
XX Example 1; Page 59-65; 84pp; English.
XX
XX This invention relates to a polynucleotide sequence encoding a
XX cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
XX PGP protein. PGP, also known as multidrug transporter, MDRL1 is a member
XX of the ABC transporter superfamily. The enzyme serves as an efflux pump
XX exporting small molecules across the cell membrane. The invention
XX includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
XX and protein, and also that of an allelic variant. The PGP polynucleotide
XX sequence is useful for increasing PGP transporter activity in a cell.
XX Antisense sequences of the cDNA are useful for inhibiting PGP transport
XX activity in a mammalian cell. They may also be used for increasing the
XX bioavailability of a drug. The present sequence represents cDNA encoding
XX cynomologous monkey P-glycoprotein. This sequence contains a 9 nucleotide
XX CC insert compared to the PGP allelic variant given in AAF86127.
XX
SQ Sequence 4195 BP; 1230 A; 801 C; 1039 G; 1125 T; 0 other;
Alignment Scores:
Pred. No.: 7.9e-191 Length: 4195

Score: 1678.00 Matches: 333
Percent Similarity: 95.38% Conservative: 18
Best Local Similarity: 90.49% Mismatches: 17
Query Match: 91.39% Indels: 0
DB: 22 Gaps: 0
US-09-672-725c-6 (1-368) x AAF86128 (1-4195)
QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaAtrg 20
Db 1780 ACGTCAGCCTTGGACACAGAAAGTGAAGCAGTGGTTCAGGTGCTCTGGATAAGCCAGA 1839
QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1840 AAAGTCGGACCACCAATTGTGATAGCTCATCGTTTGTCTACGCTTGTATGCGCAGCTC 1899
QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1900 ATCGTGGTTTCGATGATGAGTCATTGGAGAAAGAAATCATGATGAGCTCATGAA 1959
QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
Db 1960 GAGAAAGGCATTTACTCAAACTTGTCACAATGCACAGCAGAGAAATGAATGAATTA 2019
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 2020 GAAAATGCAGCTGATGAATCCAAAGTGAATGATACCTTGGAAATGCTTTCACATGAT 2079
QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
Db 2080 TCAGGATCCAGCTAATAAGAAAGATCCACTCGTAGAGTGTCCTGGATGATCACAAGC 2139
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
Db 2140 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGTATACCTCCAGTTCCCTTT 2199
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
Db 2200 TGGAGGATTATGAAGCTAAATTAACGTAGTGGCTTATTTGTTGTTGTTGTTGTTGTT 2259
QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 2260 GCCATTATAAATGGAGTCTGCACAGCATTTGCAGTAATATTTTCAAGATTATAGGG 2319
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2320 ATTTTACAAAGAAATGATGATCCGAAACAAACACAGAGAAATAGTAAGTTTTCACATA 2379
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 2380 TTGTTTCTAGTCTCTGGAAATGTTTCTTTTATTACATTTTCTCTCAGGGCTTCACATTT 2439
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2440 GCAAAAGCTGGAGAGATCCTCACCAGCGGCTCCGATACATGGTTTCCGATCCATGCTC 2499
QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 2500 AGACAGGATGTGAGCTGGTTTGTATGACCCCTAAACACACCACTGGAGCATTCGACTACCAGG 2559
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2560 CTCGCCAATGATGTCGTCGAAGTTAAAGGGGCTATAGGTTCCAGGCTTGCTATATACC 2619
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 2620 CAGATATAGCAAAATCTTTGGGACAGAGATATATATATCTTAATCTATGTTGGCAACTG 2679
QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
Db 2680 ACACGTGTTACTCTTAGCAATTTGACCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAA 2739
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340


```
XX DE Multidrug-resistance gene MDR-1.
XX KW Multidrug-resistance; MDR-1 gene; ds.
XX OS Synthetic.
XX PN W09417210-A.
XX PD 04-AUG-1994.
XX PF 25-JAN-1994; 94WO-US00748.
XX PR 27-JAN-1993; 93US-0010433.
XX PR 16-DEC-1993; 93US-0168621.
XX PA (ONCO-) ONCOR INC.
XX PI Bhatnagar SK, George AL;
XX DR WPI; 1994-264122/32.
XX PT Enzymatic amplification of target nucleic acid sequences to form
XX PT exact or modified copies - has increased fidelity and can identify
XX PT point mutations or allele(s)
XX PS Disclosure; Page 30; 69pp; English.
XX CC This gene is used as an example of a target DNA in a method for the
XX CC amplification of nucleic acid sequences in a mixture using various
XX CC DNA probes/primers. The method may be used to detect a mutation or
XX CC allele in the target, e.g. that causes a genetic disease.
XX SQ Sequence 2726 BP; 818 A; 538 C; 642 G; 728 T; 0 other;

Alignment Scores:
Pred. No.: 2.13e-190 Length: 2726
Score: 1672.00 Matches: 333
Percent Similarity: 94.84% Conservative: 16
Best Local Similarity: 90.49% Mismatches: 19
Query Match: 91.07% Indels: 0
DB: 15 Gaps: 0

US-09-672-725c-6 (1-368) x AAQ070916 (1-2726)

QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValAlaLeuAspLysAlaArg 20
DQ 176 ACCTCAGCCTTGGACACAGAAACCGAAGCAGTGGTTCAGGTGCTCTGGATAGGCCAGA 235
QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
DQ 236 AAAGGTGGGACCACCATTTGTGATAGCTCATCGTTTCTACAGTTCGTAATGCTGACGTC 295
QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
DQ 296 ATCGCTGGTTCGATGATGAGTCATGTGGAGAAAGGAATCATGATGAACATCATGAAA 355
QY 61 GluLysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
DQ 356 GAGAAAGGCATTTTACTTCAAACTGTGCACATGCACAGCAGCAGGAATGAAGTTGAATTA 415
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
DQ 416 GAAATGCAAGCTGATGAATCCAAAGTGAATTTGATCCCTTGGAAATGCTTCAAATGAT 475
QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
DQ 476 TCAAGATCCAGCTATATAGAAAAGATCAACTCGTAGGAGTGTCCTGGATCAACAAGCC 535
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
DQ 536 CAAGACAGAAAGCTAGTACCAAGAGGCTCTGGATGAAGTATACCTCCAGTTTCCTTT 595
```

```
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProThrPheValValGlyIlePheCys 160
DQ 596 TGGAGGATTATGAAGCTAAATTAACCTGAATGGCCTATTTTCTGTTGGTGTATTGTTGT 655
QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
DQ 656 GCCATTATAAATGGAGCCTGCAACCCAGCATTTGCAATAATATTTTCAAGATTATAGGG 715
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
DQ 716 GTTTTACAAAGATTGATGATCTCTGAACAAACAGCAGCAATAGTAACCTGTTTTCAC 775
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
DQ 776 TTGTTTCTAGCCCTTGGAAATTTCTTTATTACATTTTCTCTCAGGGTTTCACATTT 835
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgLysMetValPheArgSerMetLeu 240
DQ 836 GGCAAAAGCTGGAGAGATCTCCACCAAGCGCTCCGATACATGGTTTCCGATCCATGCTC 895
QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
DQ 896 AGACAGGATGTGAGTTGGTTTGATGACCTTAAACACACCATGGAGCATTTGACTACCAGG 955
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
DQ 956 CTCGCCAATGATGCTCTCAAGTTAAAGGGCTATAGTTCCAGGCTTCTGTAATTAAC 1015
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleThrGlyTrpGlnLeu 300
DQ 1016 CAGATATAGCAAACTCTGGGACAGAAATATATATCTTCTCATCTATGTTGGCAACTA 1075
QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
DQ 1076 ACATGTTTACTCTTAGCAATTTACCCATCATTTGCAATAGCAGAGTGTGTTGAAATGAA 1135
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 340
DQ 1136 ATGTTGCTGGACAGCACTGAAGATAGAAAGAACTAGAAAGGTGCTGGGAGATCGCT 1195
QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluLysPheGlu 360
DQ 1196 ACTGAAGCAATAGAAAACCTCCGAAACCGTTGTTCTTTGACTCAGGACAGAGTTTGAA 1255
QY 361 HisMetThrAlaGlnSerLeuGln 368
DQ 1256 CATATGATGCTCAGAGTTTGCAG 1279
RESULT 11
AAT43322
ID AAT43322 standard; DNA; 2726 BP.
XX
AC AAT43322;
XX
DT 01-SEP-1997 (first entry)
XX
DE Multidrug resistance gene-1.
XX
KW PCR; primer; amplify; polymerase chain reaction; ligase chain reaction;
KW LCR; human; multidrug resistance gene; MDR-1; ds.
XX
OS Synthetic.
XX
PN W09639537-A1.
XX
PD 12-DEC-1996.
XX
PF 04-JUN-1996; 96WO-US08841.
XX
PR 05-JUN-1995; 95US-0461823.
XX
XX (ONCO-) ONCOR INC.
XX
```

PI Bhatnagar SK, George AL, Nazarenko I;

XX
DR
WPT: 1997-043158/04.

PT Amplification method avoiding strand displacement by polymerase -
PT used in the detection of mutation(s) and allele(s) associated with
PT genetic disease and cancer

XX
PS
Example 1: Page 50-51: 92pp: English.

This sequence represents the multidrug resistance gene (MDR-1) amplified by the primers shown in AM43320 and AM43321. This sequence can be used as a target in the method of the invention, for enzymatically amplifying a target nucleic acid (TNA) sequence contained in a nucleic acid or mixture of nucleic acids while avoiding strand displacement by polymerase. The method comprises using three primers, one complementary to a first segment of the TNA, a second complementary to a second segment of the TNA, which is adjacent to the first primer, and a third which is similar to the first segment of the TNA, and is complementary to a portion of the first primer. The first two primers are hybridised to the TNA, and a fused amplification product is created from the TNA using these two primers. The fused product is dissociated, and hybridised to the third primer, which is then extended. The extended modified amplification product is hybridised to the first two primers, and these two primers are ligated. Each of the primers may be labelled using a different label so that the method can be used to detect the presence of a mutation or allele by detecting whether the labelled primer is contained within the fused amplification product or the extended amplification product. The method of the invention combines certain aspects of ligase chain reaction (LCR) and polymerase chain reaction (PCR), but is improved compared to LCR due to the reduced number of primers needed, and the fact that the entire TNA sequence does not need to be known.

XX sequence 2726 BP: 818 A; 538 C; 642 G; 728 T; 0 other;
 SO

Alignment Scores:		
Pred. No.:	2,13e-190	Length: 2726
Score:	1672.00	Matches: 333
Percent Similarity:	94.84%	Conservative: 16
Best Local Similarity:	90.49%	Mismatches: 19
Query Match:	91.07%	Indels: 0
DB:	18	Gaps: 0

ΠS-09-672-725C-6 (1-368) x AAT43322 (1-2726)

Qy	1	ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg	20
Db	176	ACGTGAGCTTTGGACACAGAAAACGAGCAGTGGTTTCAGTGGCTCTGGATAAGCCAGA	235
Qy	21	LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal	40
Db	236	AAAGTCGGACCCACCATTTGGTAGCTCATCGTTTGTCACAGTTCGTAAATGCTGACGTC	295
Qy	41	IleAlaGlyPheAspAspGlyValIleValIleGlyLysGlyAsnHisAspGluLeuMetLys	60
Db	296	ATCGCTGTTTTCGATGATGGATCATTTGTGGAGAAAGAAATCATCATCAATCATGAA	355
Qy	61	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu	80
Db	356	GAGAAAGCATTTACTTCAAACTTGTCCAAATGTCAGACAGCAGGAATGAAGTTCAATTA	415
Qy	81	GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysasp	100
Db	416	GAATAATGCAGCTGATGAATCCAAAAGTGAATTTGATGCCCTTGGAAATGCTTCCAATGAT	475
Qy	101	SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly	120
Db	476	TCAGATCCAGTCTCTATPAGAAAAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAAGCC	535
Qy	121	GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe	140
Db	536	CAAGACAGAAAGCTTAGTACCAAGAGAGCTCTGGATCAAAAGTAACTCTCAGTTCCCTT	595

FT /*tag= b
 FT /note= "cDNA sequence of G185V human mutant MDR-1 given
 FT in AAZ49333"
 XX

PN W09961589-A2.

XX 02-DEC-1999.

PD 27-MAY-1999; 99WO-US11825.

PF 28-MAY-1998; 98US-0086988.

PR (SUUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PA Sorrentino B, Bunting K;

PI WPI: 2000-072615/06.

DR P-PSDB; AAY59186.

XX Ex vivo expansion of hematopoietic stem cells transduced with a
 sequence encoding human multidrug resistance-1, used for bone marrow
 transplantation -

PS Claim 10; Page 68-70; 113pp; English.

XX This sequence represents cDNA encoding human wild-type
 CC multidrug resistance protein MDR-1. MDR-1 is a transmembrane
 CC efflux pump, responsible for the export of drugs from cells,
 CC particularly cancer cells. Wild-type MDR-1 shows increased
 CC resistance to ecoposide and decreased resistance to vinca
 CC alkaloids compared with a mutant form (AAY58187) where the Gly at
 CC position 185 is replaced by Val. The invention relates to transducing
 CC hematopoietic stem cells with nucleic acid encoding an MDR protein
 CC and culturing the modified cells. The modified haematopoietic stem
 CC cells are useful in bone marrow transplantation (to reconstitute
 CC hematopoietic systems in patients who have undergone chemotherapy or
 CC radiation therapy) and in ex vivo gene therapy of genetic defects in
 CC cells derived from haematopoietic stem cells, e.g., thalassemia,
 CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
 CC cells can also be used to identify factors involved in regulating
 CC proliferation and differentiation in haematopoietic stem cells.
 CC Haematopoietic stem cells that express MDR-1 will be protected against
 CC chemotherapeutic agents, so can be engrafted while the patient is
 CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
 CC provides sufficient cells to permit standard biochemical analysis.
 CC Overexpression of MDR-1 allows cytokine-driven expansion of
 CC haematopoietic stem cells by at least 10-fold compared with a maximum
 CC of 4-fold in known procedures.

XX Sequence 3860 BP; 1135 A; 746 C; 957 G; 1022 T; 0 other;

SQ Alignment Scores:

Pred. No.:	3,66e-190	Length:	3860
Score:	1672.00	Matches:	333
Percent Similarity:	94.84%	Conservative:	16
Best Local Similarity:	90.49%	Mismatches:	19
Query Match:	91.07%	Indels:	0
DB:	21	Gaps:	0

US-09-672-725c-6 (1-368) x AAZ49332 (1-3860)

QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
 |||||
 Db 1672 ACGTGACCTTGGACACAGAAAGCGAGCGTGGTTCAGGTGGCTCGGTAAGGCCAGA 1731
 QY 21 LysGlyArgThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
 |||||
 Db 1732 AAAGGTCGGACCACTTGGATGAGTCATCGTTGCTACAGTTGCTAATGCTGACGTC 1791
 QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
 |||||
 Db 1792 ATCGCTGGTTTCGATGATGAGTCATTTGGAGAAAGGAATCATGATGAATCATGAA 1851

QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
 |||||
 Db 1852 GAGAAAGGCATTACTTCAAACTTGTCACAAATGCACAGCAGCAAAATGAAGTTGAATTA 1911
 QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
 |||||
 Db 1912 GAAAATGCAGCTGATGAATCCAAAAGTGAATTTGATGCGCTTGGAAATGCTTCAAAATGAT 1971
 QY 101 SerGlySerSerLeuIleLysArgSerThrArgSerIleHisAlaProGlnGly 120
 |||||
 Db 1972 TCAAGATCCAGTCTAATAAGAAAAGATCAACTCGTAGGAGTTCGCTGGATCACAAGCC 2031
 QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProValSerPhe 140
 |||||
 Db 2032 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 2091
 QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTyrProTyrPheValValGlyIlePheCys 160
 |||||
 Db 2092 TGCAGGATTAAGAAGCTAAATTTAACTGAATGSCCTTATTTGTTGTTGGTATTTTCT 2151
 QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleLeGly 180
 |||||
 Db 2152 GCCATTATAAATGGAGGCTGCAACAGCAGATTTGCAATAATATTTTCAAAAGATTATAGG 2211
 QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
 |||||
 Db 2212 GTTTTACAGAAATGATGATCTGAAACAAACAGCAGAAATAGTAACCTGTTTTCACATA 2271
 QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
 |||||
 Db 2272 TTGTTTCTAGCCCTTGAATTTATTTCTTTATACATTTTCTTCAAGTTTTCACATTT 2331
 QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
 |||||
 Db 2332 GGCAAAGCTGGAGAGATCTCCACAGCGGCTCCGATACATGTTTCCGATCCATGCTC 2391
 QY 241 ArgGlnAspValSerTrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
 |||||
 Db 2392 AGACAGATGTGAGTTGGTTTGTATGACCCCTAAACACACCCTGGAGCATTGACTACCAAG 2451
 QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
 |||||
 Db 2452 CTCGCCAATGATGTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGTCTAATACC 2511
 QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
 |||||
 Db 2512 CAGAAATAGCAAACTCTGGGACAGGAATAATATATCTTCATCTATGGTTGGCAACTA 2571
 QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
 |||||
 Db 2572 ACACGTGTACTCTTAGCAATGTACCCATCATTTGCAATAGCAGGAGTTGTTGAATGAA 2631
 QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
 |||||
 Db 2632 ATGTTGTTCTGGACAAAGCATCTGAAGATAAGAAAGAACTAGAAAGGTGCTGGGAAGATCGCT 2691
 QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
 |||||
 Db 2692 ACTGAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTGACTCAGGACAGCAAGTTTGA 2751
 QY 361 HisMetTyrAlaGlnSerLeuGln 368
 |||||
 Db 2752 CATATGTATGTCACAGTTTGCAG 2775
 RESULT 13
 AAZ49333
 ID AAZ49333 standard; cDNA; 3860 BP.
 XX
 AC AAZ49333;
 XX
 DT 14-MAR-2000 (first entry)
 XX
 DE Human G185V mutant multidrug resistance-1 (MDR-1) cDNA.

XX Multidrug resistance: MDR-1; P-glycoprotein;
 KW transmembrane efflux pump; haematopoietic stem cell; transduction;
 KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
 KW gene therapy; gene replacement; genetic defect; thalassaemia;
 KW Gauchier's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
 KW cytokine; mutant; ds.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 XX Location/Qualifiers
 FH 1..3843
 FT /*tag= a
 FT /product= "Human G185V mutant MDR-1 protein"
 XX
 XX W09961589-A2.
 XX
 XX
 XX 02-DEC-1999.
 XX
 XX 27-MAY-1999; 99WO-US11825.
 XX
 XX 28-MAY-1998; 98US-0086988.
 XX
 XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 XX
 XX Sorrentino B, Bunting K;
 XX
 XX WPI: 2000-072615/06.
 XX P-PSDB; AAY58187.
 XX
 XX Ex vivo expansion of hematopoietic stem cells transduced with a
 XX sequence encoding human multidrug resistance-1, used for bone marrow
 XX transplantation -
 XX
 XX Example 1; Page 79-82; 113pp; English.
 XX
 XX This sequence represents cDNA encoding human G185V mutant multidrug
 XX resistance protein MDR-1, where the Gly residue at position 185
 XX of the wild-type protein (AAY58186) is replaced by Val. MDR-1 is a
 XX transmembrane efflux pump, responsible for the export of drugs from
 XX cells, particularly cancer cells. The wild-type MDR-1 shows increased
 XX resistance to etoposide and decreased resistance to vinca alkaloids
 XX compared with the G185V mutant. The invention relates to transducing
 XX haematopoietic stem cells with nucleic acid encoding an MDR protein
 XX and culturing the modified cells. The modified haematopoietic stem
 XX cells are useful in bone marrow transplantation (to reconstitute
 XX haematopoietic systems in patients who have undergone chemotherapy or
 XX radiation therapy) and in ex vivo gene therapy of genetic defects in
 XX cells derived from haematopoietic stem cells, e.g., thalassaemia,
 XX Gauchier's disease, sickle cell anaemia or leukaemia. The modified
 XX cells can also be used to identify factors involved in regulating
 XX proliferation and differentiation in haematopoietic stem cells.
 XX Haematopoietic stem cells that express MDR-1 will be protected against
 XX chemotherapeutic agents, so can be engrafted while the patient is
 XX undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
 XX provides sufficient cells to permit standard biochemical analysis.
 XX Overexpression of MDR-1 allows cytokine-driven expansion of
 XX haematopoietic stem cells by at least 10-fold compared with a maximum
 XX of 4-fold in known procedures.
 XX
 XX Sequence 3860 BP; 1135 A; 746 C; 956 G; 1023 T; 0 other;
 XX
 XX Alignment Scores:
 XX Pred. No.: 3,66e-190 Length: 3860
 XX Score: 1672.00 Matches: 333
 XX Percent Similarity: 94.84% Conservatives: 16
 XX Best Local Similarity: 90.49% Mismatches: 19
 XX Query Match: 91.07% Indels: 0
 XX DB: 21 Gaps: 0
 XX
 XX US-09-672-725C-6 (1-368) x AA249333 (1-3860)

Qy 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
 Db 1672 ACGTCAGCCCTGGACACAGAAAGCGAAGCAGTGGTTCAGGTGGCTCTGGATAAGGCCAGA 1731
 Qy 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
 Db 1732 AAAGTGGAGCCACCATTGGTATAGCTCATCGTTTGTCTACAGTTCGTAATGCTGACGTC 1791
 Qy 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
 Db 1792 ATCGCTGTTTCGATGATGGAGTCATTGTGGAGAAAGAAATCATGATGAACATCATGAA 1851
 Qy 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
 Db 1852 GAGAAAGGCATTCTTCAAACTTGTCCAAATGCAGACAGCAGGAAATGAAGTTGAAATTA 1911
 Qy 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
 Db 1912 GAAATGCAGCTGATGAATCCAAAAGTGAATTCATGCTTGGAAATCTCTTCAATGAT 1971
 Qy 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
 Db 1972 TCAAGATCCAGCTAATAAGAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAGCC 2031
 Qy 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluValProValSerPhe 140
 Db 2032 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTCCCTT 2091
 Qy 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
 Db 2092 TGGAGGATTAAGAAGCTAAATTAACGAATGGCCCTATTATTTGTGTGGTGATTGTGT 2151
 Qy 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
 Db 2152 GCCATTATAAATGGAGGCTGCAACAGCAGCATTTGCAATAAATATTTTCAAGATATAGG 2211
 Qy 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
 Db 2212 GTTTTACAAAGAAATTTGATGATCCTGAAACAAACAGCAGAAATAGTAATCTGTTTCACTA 2271
 Qy 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
 Db 2272 TTGTTTCAGCCCTTGGAAATTTATTTCTTTTATACATTTTCTTCAAGGTTTTCACATT 2331
 Qy 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
 Db 2332 GCAGAGCTGGAGAGATCCTCACCAGGCGCTCCGATACATGGTTTCCGATCCATGCTC 2391
 Qy 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
 Db 2392 ACACAGGATGTCAGTTGGTTTCATGACCCCTAAACACACACTGGAGCATTGACTACCAGG 2451
 Qy 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
 Db 2452 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGGTTCCAGGCTTGTGTAATACC 2511
 Qy 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
 Db 2512 CAGAATATAGCAAACTCTGGGACAGGAATAATATATATCTTCATCTATGTTGGCAACTA 2571
 Qy 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
 Db 2572 ACACGTGTACTCTAGCAATTTACCCATCATTTGCAATAGCAGGAGTGTGTTGAAATGAA 2631
 Qy 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyValGlyLysIleAla 340
 Db 2632 ATGTTGCTGTCGACAGCAGCTGAAAGATAAGAAAGAACTAGAAGGTGCTGGGAGATCGCT 2691
 Qy 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGlnLysPheGlu 360
 Db 2692 ACTGAAGCAATAGAAAACCTCCGACCGTGTGTTCTTCTGACTCAGGAGCAGAGTTTGA 2751
 Qy 361 HisMetTyrAlaGlnSerLeuGln 368

Thu Nov 7 09:13:35 2002

Db 2752 CATATGTATGCTCAGAGTTTGCAG 2775

RESULT 15

ABA94366

ID ABA94366 standard; DNA; 3860 BP.

XX ABA94366;

XX 26-MAR-2002 (first entry)

XX Human BCRP DNA related seq Id No. 3.

XX Stem cell; ATP transport protein; ATP-binding cassette; antiparkinsonian;
KW hepatotropic; neurodegenerative; cytostatic; antianemic; muscular; BCRP;
KW cardiant; gene therapy; ds.

XX Homo sapiens.

OS Key Location/Qualifiers

FT 1..3843

FT /*tag= a

XX W0200192877-A2.

XX 06-DEC-2001.

XX 30-MAY-2001; 2001W0-US17459.

XX 31-MAY-2000; 2000US-0584586.

XX 29-MAY-2001; 2001US-0866866.

XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX Sorrentino B, Schuetz J;

XX WPI; 2002-114368/15.

XX P-PSDB; ABB07267.

XX Identifying a stem cell, for treating e.g., muscular dystrophy,
PT myocardial infarction, Parkinson's disease, or neurodegenerative
PT disorders, comprises detecting the expression of an ATP transport
PT protein (BCRP) by a cell -

XX Disclosure; Page 59-60; 87pp; English.

XX The invention provides a method of identifying and/or isolating a stem
CC cell that involves detecting the expression of an ATP transport protein
CC containing a conserved ATP-binding cassette (BCRP) by a cell in a sample
CC comprising stem cells. The isolated stem cells may be used in the
CC treatment of diseases such as muscular dystrophy, degenerative liver
CC disorder, myocardial infarction, Parkinson's disease, degenerative
CC disorders of the brain, and for tissue regeneration or replacement.
CC Haematopoietic cells can be used in bone marrow transplants (e.g., for
CC treatment of leukemia) and for ex vivo gene therapy for treating blood
CC diseases such as sickle cell anemia and thalassemia. The stem cells can
CC also be used as cell targets in gene therapy protocols. The present
CC sequence represents a sequence related to the BCRP for which no relevant
CC information has been provided in the specification.

XX SQ Sequence 3860 BP; 1135 A; 746 C; 956 G; 1023 T; 0 other;

Alignment Scores:

Pred. No.: 3,65e-190 Length: 3860
Score: 1672.00 Matches: 333
Percent Similarity: 94.84% Conservative: 16
Best Local Similarity: 90.49% Mismatches: 19
Query Match: 91.07% Indels: 0
DB: 24 Gaps: 0

US-09-672-725C-6 (1-368) x ABA94366 (1-3860)

QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArg 20

|||||

|||||

|||||

|||||

|||||

|||||

|||||

Db 1672 AGTCAGCCTTGGACACAGAAAGCCAGCAGTGGTTCAGGTGGCTCTGGATAAGCCAGA 1731

QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40

|||||

Db 1732 AAAGTCGGACCACCATTGTGATAGTCATCGTTTGTCTACAGTTCGTAATGCTGACGTC 1791

QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60

|||||

Db 1792 ATCGCTGGTTTCGATGGAGTCAATTCGAGAAAGGAAATCATGATGAAGTTCATGAA 1851

QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80

|||||

Db 1852 GAGAAGGCATTTACTTCAAACTGTCAATGCAGACAGCAGGAAATGAAGTTGAATTA 1911

QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100

|||||

Db 1912 GAAATGCGAGCTGATCAATCCAAAAGTGAATTCATGCTTGGAAATGCTCTCAATGAT 1971

QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120

|||||

Db 1972 TCAAGATCCAGTCTAATAAGAAAAAGATCAACTCTGAGAGTGTCCGTGGATCACAGCC 2031

QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140

|||||

Db 2032 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 2091

QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160

|||||

Db 2092 TGGAGGATTATGAAGCTAAATTTAACTGAATGGCCCTATTTTGTGGTGTATTTTGT 2151

QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180

|||||

Db 2152 GCCATTATAATGGAGGCTGCACACAGCATTGCAATAATATTTTCAAGATATATAGGG 2211

QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200

|||||

Db 2212 GTTTTACAAAGAAATGATGATCCTGAAACAAACACAGAAATAGTAATCTTTTCACTA 2271

QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220

|||||

Db 2272 TTGTTTCTAGCCCTTGGAAATATTTCTTTTATACATTTTCTTCAAGTTTTCATTT 2331

QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240

|||||

Db 2332 GCCAAGCTGGAGAGATCCTCACCAAGCGCTCCGATACATGTTTCCGATCCATGCTC 2391

QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260

|||||

Db 2392 AGACAGGATGTCAGTTGGTGTGATGACCCCTAAAAACACCACTGGAGCATTGACTACCAGG 2451

QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280

|||||

Db 2452 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTGCTGTATTACC 2511

QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300

|||||

Db 2512 CAGAATATAGCAAACTTGGGACAGCAATAATATATATCTTCTCATCTATGTTGGCACTA 2571

QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320

|||||

Db 2572 ACACCTGTTACTCTTAGCAATTTACCCATTCATTCGAATAGCAGGAGTGTGTTGAAATGAA 2631

QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyValaGlyLysIleAla 340

|||||

Db 2632 ATGTTGCTGGACACAGCATCGAAAGATAAGAAAGTGTGTTGGAAGATCGCT 2691

QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360

|||||

Db 2692 ACTGAAGCAATAGAAAACTTCCGAACCGTGTGTTCTTTGACTCAGGACAGCAAGTTTGA 2751

QY 361 HisMetTyrAlaGlnSerLeuGln 368

|||||

Db 2752 CATATGTATGCTCAGAGTTTGCAG 2775

Search completed: November 6, 2002, 19:11:59
Job time : 131.655 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 18:39:50 ; Search time 16.8217 Seconds

(without alignments)
5373.600 Million cell updates/sec

Title: US-09-672-725C-6

Perfect score: 1836

Sequence: 1 TSALDTESEAVVQALDKAR.....TVVSLTQEQKFEHMYAQSILQ 368

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09672725/runat_04112002_124339_17887/app.query.fasta_1.6435
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09672725.@cgn2_1_134.@runat_04112002_124339_17887 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	1672	91.1	2726	1	US-08-461-823-1 Sequence 1, Appli
2	1672	91.1	4264	2	US-08-784-649A-1 Sequence 1, Appli
3	1672	91.1	4264	2	US-08-784-649A-5 Sequence 5, Appli
4	1672	91.1	4646	1	US-08-181-471-2 Sequence 2, Appli
5	1672	91.1	4669	6	5206352-3 Patent No. 5206352
6	1672	91.1	6505	2	US-08-793-610-5 Sequence 5, Appli
7	1672	91.1	9318	2	US-08-793-610-6 Sequence 6, Appli
8	1653	89.0	4669	2	US-08-583-276-18 Sequence 18, Appli
9	1644	89.5	4669	2	US-08-752-447-1 Sequence 1, Appli
10	1426.5	77.7	4233	3	US-09-120-513-1 Sequence 1, Appli
11	1426.5	77.7	4233	3	US-09-450-105-1 Sequence 1, Appli
12	591	32.2	4002	2	US-08-996-545-1 Sequence 1, Appli

13	591	32.2	4002	2	US-08-996-545-3 Sequence 3, Appli
14	591	32.2	4002	4	US-09-328-320-1 Sequence 1, Appli
15	591	32.2	4002	4	US-09-328-320-3 Sequence 3, Appli
16	588.5	32.1	4047	2	US-08-612-734B-1 Sequence 1, Appli
17	588.5	32.1	4800	2	US-08-612-734B-3 Sequence 3, Appli
18	544.5	29.7	3924	1	US-08-395-246C-1 Sequence 1, Appli
19	521.5	28.4	4224	1	US-08-612-521-1 Sequence 1, Appli
20	433	23.6	3909	1	US-08-232-537-1 Sequence 1, Appli
21	406	22.1	3924	2	US-08-996-644-3 Sequence 3, Appli
22	406	22.1	3924	3	US-09-352-552-3 Sequence 3, Appli
23	406	22.1	3927	2	US-08-996-644-1 Sequence 1, Appli
24	406	22.1	3927	3	US-09-352-552-1 Sequence 1, Appli
25	388.5	21.2	6143	1	US-08-612-521-3 Sequence 3, Appli
26	194.5	10.6	4781	2	US-09-001-273-1 Sequence 1, Appli
27	194.5	10.6	4781	4	US-08-843-459A-1 Sequence 1, Appli
28	194.5	10.6	4847	3	US-09-061-400-1 Sequence 1, Appli
29	187	10.2	1201	3	US-08-961-083-189 Sequence 189, App
30	185.5	10.1	4403765	4	US-09-103-840A-2 Sequence 2, Appli
31	181	9.9	535	4	US-08-858-207A-230 Sequence 230, App
32	174.5	9.5	7721	3	US-08-772-270A-14 Sequence 14, Appli
33	172.5	9.4	1959	4	US-09-061-764A-4 Sequence 4, Appli
34	172.5	9.4	2061	4	US-09-061-764A-17 Sequence 17, Appli
35	171.5	9.3	3576	4	US-08-976-259-79 Sequence 79, Appli
36	167	9.1	5120	3	US-08-772-270A-6 Sequence 6, Appli
37	166.5	9.1	8370	2	US-08-488-706-1 Sequence 1, Appli
38	162.5	8.9	28804	2	US-08-592-874-1 Sequence 1, Appli
39	162.5	8.9	28804	3	US-09-096-942-2 Sequence 2, Appli
40	162.5	8.9	28804	3	US-09-096-867-2 Sequence 2, Appli
41	160	8.7	4394	2	US-08-750-152A-1 Sequence 1, Appli
42	157.5	8.6	5889	1	US-08-463-092B-5 Sequence 5, Appli
43	157.5	8.6	5889	2	US-08-462-109A-5 Sequence 5, Appli
44	157.5	8.6	5889	2	US-08-460-907B-5 Sequence 5, Appli
45	157.5	8.6	5889	3	US-08-463-179A-5 Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-08-461-823-1

; Sequence 1, Application US/08461823

; Patent No. 5593840

; GENERAL INFORMATION:

; APPLICANT: Bhatnagar, Satish K.

; APPLICANT: George Jr., Albert L.

; APPLICANT: Nazarenko, Irina

; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACID SEQUENCES

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OncorPharm, Inc.

; STREET: 200 Perry Parkway

; CITY: Gaithersburg

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20877

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/461,823

; FILING DATE: 05-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/168,621

; FILING DATE: 16-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/010,433

; FILING DATE: 27-JAN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Karta, Glenn E.

; REGISTRATION NUMBER: 30,649

; REFERENCE/DOCKET NUMBER: PA-0012 CIP 2

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301 527-2058
 TELEFAX: 301 208-6997
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2726 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-461-823-1

Alignment Scores:
 Pred. No.: 3,74e-206 Length: 2726
 Score: 1672.00 Matches: 333
 Percent Similarity: 94.84% Conservative: 16
 Best Local Similarity: 90.49% Mismatches: 19
 Query Match: 91.07% Indels: 0
 DB: 1 Gaps: 0

US-09-672-725c-6 (1-368) x US-08-461-823-1 (1-2726)

QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
 DB 176 ACCTCAGCCTTGGACACAGAAAGGAGCAGTGGTTCAGTGGCTCTGGATAAGGCCAGA 235
 QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
 DB 236 AAAGGTGGACACCACTTGTGATAGCTCATCGTTTCTACAGTTCGTAAATGCTGACGTC 295
 QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyValAsnHisAspGluLeuMetLys 60
 DB 296 ATCCTGTTTCATGATGAGGAGTCATGTGGAAAGAAATCATGATCACTCATGAAA 355
 QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
 DB 356 GAGAAAGGCATTACTTCAAACTGTGCACAAATGCACAGCAGCAGCAATGAATGTAATTA 415
 QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
 DB 416 GAAATGCAGCTGATGAATCCAAAGTGAAATGATGCTTGGAAATGCTTCAAAATGAT 475
 QY 101 SerGlySerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 120
 DB 476 TCAAGATCCAGCTCTAATAAGAAAGAGTCAACTCGTAGGAGTGTCCGTGGATCACAAGCC 535
 QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluValProProValSerPhe 140
 DB 536 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGTATACCTCCAGTTCCTTT 595
 QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
 DB 596 TGGAGGATATGAGACCTTAATTAACCTGATGCTTATTTGTTGTTGTTGTTGTTGTTGTT 655
 QY 161 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleLeuPheSerArgIleGly 180
 DB 656 GCCATTATAAATGAGCGCTGCAACACAGCATTTGCAATAATATTTTCAAAAGATTATAGGG 715
 QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
 DB 716 GTTTTCAAGAATGTGATGATCCGAAACAAACAGCAGCAATAGTAACTTTCTTTTCACTA 775
 QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
 DB 776 TTGTTTCTAGCCCTTGGAAATATTCTTTTATTACATTTTTCCTTCAGGTTTTCACATTT 835
 QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
 DB 836 GCACAAAGCTGGAGATCCTCACCAGCGCTCCGATACATGGTTTCCGATCCATGCTC 895
 QY 241 ArgGlnAspValSerTrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260

DB 896 AGACAGGATGTGAGTTGGTTTGTATGACCCAAAACACCACTGGAGCATTTGACTACGAG 955
 QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
 DB 956 CTCGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGTGTAATTACC 1015
 QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
 DB 1016 CAGAAATATAGCAATCTTGGACAGAGGAATAATTATATCTTCTATGTTGGCAACTA 1075
 QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
 DB 1076 ACATGTTTACTCTTAGCAATTTGATCCCATCATTCATATAGCAGGAGTTGTTGAATGAAA 1135
 QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
 DB 1136 ATGTTGCTGACAGCACTGAAAGATAAGAAAGAACTAGAAAGTGTCTGGAGATCGCT 1195
 QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
 DB 1196 ACTGAAGCAATAGAAACTTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAGTTTGAA 1255
 QY 361 HisMetTyrAlaGlnSerLeuGln 368
 DB 1256 CATATGTATGCTCAGAGTTTGCAG 1279
 RESULT 2
 US-08-784-649A-1
 ; Sequence 1, Application US/08784649A
 ; Patent No. 5830697
 ; GENERAL INFORMATION:
 ; APPLICANT: Sikic, Branimir I
 ; APPLICANT: Chen, Gang
 ; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
 ; TITLE OF INVENTION: CYCLOSPORIN MODULATION
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 2200 Sand Hill Road
 ; CITY: Menlo Park
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/784,649A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sherwood, Pamela J
 ; REGISTRATION NUMBER: Reg. No. 5830697 36,677
 ; REFERENCE/DOCKET NUMBER: 06037/007001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-322-5070
 ; TELEFAX: 415-854-0875
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4264 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-08-784-649A-1

Alignment Scores:
 Pred. No.: 7,97e-206 Length: 4264
 Score: 1672.00 Matches: 333
 Percent Similarity: 94.84% Conservative: 16

Thu Nov 7 09:13:36 2002

Qy	61	GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu	80
Db	1990	GAGAAAGGCAATTTACTTCAAACTTGTCAAAATGCAGACAGCAGGAAATGAAGTTGAATTA	2049
Qy	81	GluAsnAlaThrClyLysSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp	100
Db	2050	GAAAAATGCAGCTGATGAATCCAAAAAGTGAATGATGCTTGGAAATGTCTTCAAAATGAT	2109
Qy	101	SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly	120
Db	2110	TCAAGATCCAGCTCTAATAGAAAAGATCAACTCGTAGGAGTGCCGTGGATCACAAGCC	2169
Qy	121	GlnAspArgLysLeuGlyThrLysGluAspLysLeuAsnValProProValSerPhe	140
Db	2170	CAAGACAGAAAGCTTATAGTACCAAGAGGCTCTGGATGAAAGATATACCTCCAGATTTCTCTT	2229
Qy	141	TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys	160
Db	2230	TGAGGAGTATGAAAGCTAAATTTAACTGAATGGCCCTATTTTGTGTGGTGATATTTTGT	2289
Qy	161	AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly	180
Db	2290	GCCATTATAAATGGAGCCCTGCAACAGCATTGTGCAATAATATTTTCAAAAGATTATAGGG	2349
Qy	181	IlePheThrArgAspGluAspProGlnThrLysArgGlnIleSerAsnMetPheSerVal	200
Db	2350	GTTTTTCACAAAGATTGATCTGTAACAAACACAGAAATAGTAACTTGTGTTCACATA	2409
Qy	201	LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe	220
Db	2410	TTGTTTCTAGCCCTTGGAAATATTCTTTTATTACATTTTTTCCCTCAGGGTTTCACATTT	2469
Qy	221	GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu	240
Db	2470	GGCAAAAGCTGGAGAGATCTCCACAAAGGGCTCCGATACATGGTTTCCGATCCATGCTC	2529
Qy	241	ArgGlnAspValSerTrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg	260
Db	2530	AGACAGGATGTGATTGGTTTGATGACCTCAAAACACCACCTGGAGCATTTGACTACCCAGG	2589
Qy	261	LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr	280
Db	2590	CTGCCCAATGATCTGCTCAAGTTAAAGGGGCTATAGTTCCAGGCTTGCTGTAAATTACC	2649
Qy	281	GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu	300
Db	2650	CAGATATACAAATCTTGGCAGAGAAATATATATCTTCATCTATGGTTGGCAACTA	2709
Qy	301	ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys	320
Db	2710	ACACTGTTACTCTTAGCAATTTGACCCATCATTTGCAATGACGAGGATGTTGTAATGAAA	2769
Qy	321	MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGlyAlaGlyLysIleAla	340
Db	2770	ATGTTGTCTGGACAGCCTGGAAGATAAGAAGAACTAGAAAGTGCTGGGAAGATCGCT	2829
Qy	341	ThrGluAlaIleGluAsnPheArgThrValSerLeuThrGlnGluGlnLysPheGlu	360
Db	2830	ACTGAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAAAGTTTGA	2889
Qy	361	HisMetTyrAlaGlnSerLeuGln	368
Db	2890	CATATGTATGCTCAGAGTTTGGCAG	2913
RESULT 4			
US-08-181-471-2			
; Sequence 2, Application US/08181471			
; Patent No. 5641508			
; GENERAL INFORMATION:			
; APPLICANT: Li, Lingna			
; * APPLICANT: Lishko, Valeryi K.			
; TITLE/OF INVENTION: METHOD FOR DELIVERING BENEFICIAL			
; TITLE OF INVENTION: COMPOSITIONS TO HAIR FOLLICLES			

```
Db 2396 TCAAGATCCAGTCTAATAAGAAAAAGATCAACTCGTAGGAGTGTCCGTGGATCACAAGCC 2455
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
Db 2456 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTCCCTTT 2515
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
Db 2516 TGGAGGATATGAAGCTAAATTAACATGAATGGCCTTATTTGTGTGGTGTATTTGT 2575
QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 2576 GCATTAATAATGGAGGCTCGCAACAGCATTTGCAATAAATATTTCAAGATATTATAGG 2635
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2636 GTTTTACAGAATGATGATCCTGAACAAACAGCACAGATAGTAAGTCTGTTTCACTA 2695
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPheLeuGlnGlyPheThrPhe 220
Db 2696 TTGTTTCTAGCCCTTGGAAATATTTCTTTTATACATTTTCTTCAGGGTTTCACATTT 2755
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2756 GGCAAAGCTGGAGAGATCCTCACCAGCGCTCCGATACATGGTTTCCGATCCATGCTC 2815
QY 241 ArgGlnAspValSerTrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 2816 AGACAGGATGAGTGTGTTGTATGACCTAAAACACCACCTGGAGCATTTGACTACCAGG 2875
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2876 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCCAGSCTTGCTGTAATTACC 2935
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 2936 CAGAAATAGCAAACTCTGGACAGGAATAATTATATCTCATCTATGTTGGCAACTA 2995
QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
Db 2996 ACACGTGTACTCTTAGCAATTTGACCCATCATTCATATGACAGAGTGTGTGAATGAAA 3055
QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 3056 ATGTTGCTGGCAAGCACTCAAAAGATAAGAAAGAACTAGAAAGTGTCTGGGAAGTCGCT 3115
QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluLysPheGlu 360
Db 3116 ACTGAAGCAATAGAAAACCTCCGAACCGCTGTTTCTTTGACTCAGAGCAGAACTTTGAA 3175
QY 361 HisMetTyrAlaGlnSerLeuGln 368
Db 3176 CATATGATGCTCAGAGTTTGCAG 3199
```

RESULT 5

```
5206352-3
; Patent No. 5206352
; APPLICANT: Roninson, Igor B.; Pastan Ira H.; Gottesman,
; Michael M.
; TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
; SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/622,836
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 892,575
; FILING DATE: 01-AUG-1986
; APPLICATION NUMBER: 845,610
; FILING DATE: 28-MAR-1986
; SEQ ID NO:3
; LENGTH: 4669
5206352-3
```

```
Alignment Scores:
Pred. No.: 9,28e-206 Length: 4669
Score: 1672.00 Matches: 333
Percent Similarity: 94.84% Conservative: 16
Best Local Similarity: 90.49% Mismatches: 19
Query Match: 91.07% Indels: 0
DB: 6 Gaps: 0
```

US-09-672-725c-6 (1-368) x 5206352-3 (1-4669)

```
QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
Db 2096 ACGTGACGCTTGGACACAGAAAGCGAGCAGTGTTCAGGTGCTCTGGATAGGCCAGA 2155
QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 2156 AAAGGTCGGACCACTTGTGATAGCTCATCGTTGTCTACAGTTCGTATGCTGACGTC 2215
QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 2216 ATCGCTGCTTTCGATGATGGAGTCATTTGGAGAAAGGAATCATGATGAACCTCATGAA 2275
QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
Db 2276 GAGAAAGGCACTTTACTCAAACTTGTCAACATGCACAGCAGCAAGAAATGAAGTTGAATTA 2335
QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuMetSerProLysAsp 100
Db 2336 GAAATTCAGCTGATGAATCCAAAAGTGAATTTGATGCCTTGGAAATGCTCTTCAAAATGAT 2395
QY 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
Db 2396 TCAAGATCCAGTCTAATAAGAAAAGATCAACTCGTAGGAGTCTCCGTGGATCACAAGCC 2455
QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
Db 2456 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGTATACCTCCAGTTCCCTTT 2515
QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
Db 2516 TGGAGGATATGAAGCTAAATTTAACTGAATGGCCTTATTTGTTGTTGTTATTTGT 2575
QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 2576 GCCATTAATAAGTGGAGGCTCGCAACCCAGCATTTGCAATATATTTTCAAGATATTAGGG 2635
QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2636 GTTTTACAGAATTTGATGATCCTGAACAAACAGCACAGCAAGTATAGTATTTTTCACCTA 2695
QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 2696 TTGTTTCTAGCCCTTGGAAATATTTCTTTTATACATTTTCTTCAGGGTTTCACATTT 2755
QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2756 GGCAAAGCTGGAGAGATCCTCACCAGCGCTCCGATACATGGTTTCCGATCCATGCTC 2815
QY 241 ArgGlnAspValSerTrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 2816 AGACAGGATGATGATGTTGGTTTGTGATGACCTTAAAACACCACCTGGAGCATTTGACTACCAGG 2875
QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2876 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTTCCAGGCTTCTGCTGTAATTACC 2935
QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 2936 CAGAAATAGCAAACTCTGGGACAGGAATAATTATATCTCATCTATGTTGGCAACTA 2995
QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
```

Db 2996 ACAGTGTACTCTAGCAATGTACCATCATTTGCAATAGCAGGAGCTGTGTGAATGAAA 3055
Qy 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 3056 ATGTTGCTGTGGACAGCAGCTGAAAGATAAGAAAGAACTAGAAAGGTCTGGGAAGATCGCT 3115
Qy 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
Db 3116 ACTGAACAATAGAAACTTCCGACCGTGTCTTTGACTCAGGAGCAGAAGTTTGAA 3175
Qy 361 HisMetTyrAlaGlnSerLeuGln 368
Db 3176 CATATGATGCTCAGAGCTTTGCAG 3199

RESULT 6
US-08-793-610-5
; Sequence 5, Application US/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARBERS, Carol
; APPLICANT: OSTERTAG, Wolfram
; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
; TITLE OF INVENTION: FOR GENE TRANSFER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,610
; FILING DATE: 07-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 31 973.8
; FILING DATE: 08-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 195 03 952.1
; FILING DATE: 07-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/03175
; FILING DATE: 10-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Beriman, Richard J.
; REGISTRATION NUMBER: 39,105
; REFERENCE/DOCKET NUMBER: P1614-7007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6505 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
US-08-793-610-5

Alignment Scores:
Pred. No.: 1,63e-205 Length: 6505
Score: 1672.00 Matches: 333
Percent Similarity: 94.84% Conservative: 16
Best Local Similarity: 90.49% Mismatches: 19
Query Match: 91.07% Indels: 0
DB: 2 Gaps: 0

US-09-672-725c-6 (1-368) x US-08-793-610-5 (1-6505)
Qy 1 ThrSerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArg 20
Db 3488 ACGTCAGCCTTGGACACAGAAGACGACAGTGGTTCAGGTGGCTCTGGATAAGGCCAGA 3547
Qy 21 LysGlyArgThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 3548 AAGAGTCGACACCACCATTTGTGATAGTCATCGTTTGTACAGTTCGTAATGCTGACGTC 3607
Qy 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 3608 ATCGCTGGTTTCGATGATGGAGTCATTGTGGAGAAAGAAATCATGATGAATCATGAAA 3667
Qy 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
Db 3668 GAGAAAGGCATTTACTTCAAACTTGTACATGCAGACAGCAGGAAATGAAGTTGAATTA 3727
Qy 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 3728 GAAATCGAGCTGATGAATCCAAAAGTGAATTTGATGCTTTGGAATCTCTTCAATGAT 3787
Qy 101 SerGlySerSerLeuIleLysArgArgSerThrArgArgSerIleHisAlaProGlnGly 120
Db 3788 TCAAGATCCAGTCTAATAAGAAAAAGATCACTCGTAGAGTGTCCGTGGATCACAAGCC 3847
Qy 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
Db 3848 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAGATATACCTCCAGTTTCCTTT 3907
Qy 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
Db 3908 TGGAGGATTAATGAAGCTAAATTTAACTGAATGGCCCTATTTTGTGTGGTGTATTTGT 3967
Qy 161 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 3968 GCCATTATAATGGAGGCTGCACACGATTTGCAATAATATTTTCAAGATATATAGG 4027
Qy 181 IlePheThrArgAspLeuAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 4028 GTTTTTTACAAGAATTGATGATCTCTGAAACAAAACGACAGAAATAGTAACCTGTGTTTCACTA 4087
Qy 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 4088 TTGTTTCTAGCCCTTGGAAATTAATTTCTTTTATACATTTTTCCTTCAGGGTTTCACATT 4147
Qy 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 4148 GGCAAAGCTGGAGAGATCCTCACCAAGCGCTCCGATACATGGTTTTCGATCCATGCTC 4207
Qy 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 4208 AGACAGGATGTGAGTTGGTTTGTATGACCCCTAAAAACACACCTGGAGCATTTGACTACCA 4267
Qy 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 4268 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTCTGTAATTACC 4327
Qy 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 4328 CAGAATATAGCAAACTCTGGGACAGGAATAATATATATCTTTCATCTATGGTTGGCAACTA 4387
Qy 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
Db 4388 ACACCTGTACTCTTAGCAATTTACCCATCATTTGCAATAGCAGGAGTGTGTGAAATGAAA 4447
Qy 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 4448 ATGTTCTCTGGACACCACTGAAAGATAAGAAAGAACTAGAAAGGTCTGGGAAGATCGCT 4507
Qy 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
Db 4508 ACTGAAGCAATAAGAAACTTCCGACCGTGTGTTTCTTTGACTCAGGAGCAGAAGTTTGAA 4567

QY 361 HisMetTyrAlaGlnSerLeuGln 368
|||||
Db 4568 CATATGTATGCTCAGAGTTTCAG 4591

RESULT 7
US-08-793-610-6
; Sequence 6, Application US/08793610
; Patent No. 5858744
; GENERAL INFORMATION:
; APPLICANT: BAUM, Christopher
; APPLICANT: STOCKING-HARBERS, Carol
; APPLICANT: OSTERTAG, Wolfram
; TITLE OF INVENTION: RETROVIRAL VECTOR HYBRIDS AND THE USE THEREOF
; TITLE OF INVENTION: FOR GENE TRANSFER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikolaide, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,610
; FILING DATE: 07-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 31 973.8
; FILING DATE: 08-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 195 03 952.1
; FILING DATE: 07-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/03175
; FILING DATE: 10-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Beiman, Richard J.
; REGISTRATION NUMBER: 39,105
; REFERENCE/DOCKET NUMBER: P1614-7007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9318 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
US-08-793-610-6

Alignment Scores:
Pred. No.: 2,98e-205 Length: 9318
Score: 1672.00 Matches: 333
Percent Similarity: 94.84% Conservatives: 16
Best Local Similarity: 90.49% Mismatches: 19
Query Match: 91.07% Indels: 0
DB: 2 Gaps: 0

US-09-672-725c-6 (1-368) x US-08-793-610-6 (1-9318)

QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
|||||
Db 3447 ACCTCAGCCTTGGACACAGAAAGCGAGTGGTTTCAGGTGGCTCTGGTAAGGCCAGA 3506

QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
|||||
Db 3507 AAAGGTGGACCAACCAATTTGGATAGCTATCGTTTCTGTACAGTTCTGTAATGTCAGCTC 3566

QY 41 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
|||||
Db 3567 ATCGTGGTTTCGATGAGGAGTCATTTGGAGAAAGAAATCATGATGAACATCATGAA 3626

QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
|||||
Db 3627 GAGAAAGGCATTTACITCAAACTTGTCACAAATGCACAGCAGCAAGAAATGAAGTTGAATTA 3686

QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
|||||
Db 3687 GAAATGCAGCTGATGAATCCAAAGTGAATTCATGCGCTTGGAAATGCTCTCAATGAT 3746

QY 101 SerGlySerSerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 120
|||||
Db 3747 TCAAGATCCAGTCTAATAAGAAAAAGATCAACTCGTAGGAGTGTCCGTGGATCACCAGCC 3806

QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
|||||
Db 3807 CAAGACAGAAAGCTTAGTACCAAGAGGCTCTGGATGAAAGTATACCTCCAGTTCTCTT 3866

QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
|||||
Db 3867 TGGAGGATTATGAAGCTAAATTTAACTGAATGGCTTATTTGTTGTTGCTGATTTTGT 3926

QY 161 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
|||||
Db 3927 GCCATTATAATGGAGGCTGCAACAGCATTTGCAATAATATTTTCAAGATTTATAGG 3986

QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
|||||
Db 3987 GTTTTACAAAGATTGATGATCCTGAAACAAACAGACAGCAATAGTAACCTGTTTCACTA 4046

QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
|||||
Db 4047 TTGTTTCTAGCCCTTGGAAATTTCTTTTATTACATTTTCTCTTCAGGTTTTCACATTT 4106

QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
|||||
Db 4107 GGCAAAGCTGGAGAGATCTCCAAAGCGCTCCGATACATGTTTCCGATCCATGCTC 4166

QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
|||||
Db 4167 AGACAGGATGTGAGTTGGTTTGTGATGACCTCAAAACACACCTGGAGCATTTGACTACCAG 4226

QY 261 LeuAlaAsnAspAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
|||||
Db 4227 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGGTTCCAGGCTTGTCTAATTACC 4286

QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
|||||
Db 4287 CAGATATAGCAAACTTGGGACAGAAATATATATCTTCATCTATGTTGGCAACTA 4346

QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
|||||
Db 4347 ACACGTGTTACTCTTAGCAATTTACCCATCATTTGCAATAGCAGGAGTTGTTGAAATGAA 4406

QY 321 MetLeuSerGlyGlnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
|||||
Db 4407 ATGTTGCTCTGGACAAAGCACTGAAAGATAAGAAAGAACTAGAAGGTGCTGGGAAGATCGCT 4466

QY 341 ThrGluAlaIleGluAsnPheArgThrValSerLeuThrGlnGluGlnLysPheGlu 360
|||||
Db 4467 ACTGAAGCAATAGAAACCTCCGAACCGTTGTTCTTTGACTCAGGAGCAGAAAGTTTGA 4526

QY 361 HisMetTyrAlaGlnSerLeuGln 368
|||||
Db 4527 CATATGTATGCTCAGAGTTTCAG 4550

RESULT 8

US-08-583-276-18
; Sequence 18, Application US/08583276
; Patent No. 5837536

GENERAL INFORMATION:
APPLICANT: McDonagh, Kevin T.
APPLICANT: Nienhuis, Arthur
APPLICANT: Tolstoshev, Paul
TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
TITLE OF INVENTION: SELECTION OF CELLS TRANSDUCED WITH SUCH GENES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
ADDRESSEE: Cecchi & Stewart
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: DW4.V2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/583,276
FILING DATE: 05-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/332,444
FILING DATE: 31-OCT-1994
APPLICATION NUMBER: 07/887,712
FILING DATE: 22-MAY-1992
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 4669 bases
TYPE: nucleic acid
STRANDEDNESS: singular
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: Genomic DNA
US-08-583-276-18
Alignment Scores:
Pred. No.: 2,696-203 Length: 4669
Score: 1653.00 Matches: 330
Percent Similarity: 94.29% Conservative: 17
Best Local Similarity: 89.67% Mismatches: 21
Query Match: 90.03% Indels: 0
Gaps: 0
DB: 2
US-09-672-725C-6 (1-368) x US-08-583-276-18 (1-4669)
Qy 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
Db 2096 ACGTCAGCCTGGACACAGAAAGGAGGAGTGGTTCAGGTGGCTCTGGATAAGGCCAGA 2155
Qy 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 2156 AAAGGTGGACACCATTTGATAGTCATCGTTTGTCTACGTTCTGTAATGTCAGCTC 2215
Qy 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 2216 ATCCGTGTTTCGATGATGAGTCACTTGTGGAGAAAGAAATCATGATGATCATGAAA 2275
Qy 61 GluLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
Db 2276 GAGAAAGGCATTTACTTCAAACTTGTACAAATGCAGACAGCAGGAAATGAAGTTGAATTA 2335
Qy 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 2336 GAAATGCGAGCTGATGAATCCAAAGTGAAATGATGCTTGGAAATGCTTCAATATGAT 2395
Qy 101 SerGlySerLeuIleLysArgSerThrArgArgSerIleHisAlaProGlnGly 120
Db 2396 TCAAGATCACGCTCTAATAAGAAAGATCACTCGTAGGATGTCCTGGATCACACACC 2455

Qy 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
Db 2456 CAAGACAGAAAGCTTAGTACCAGAGAGCTCTGGATGAAAGATATACCTCCAGTTTCCTTT 2515
Qy 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProTyrPheValValGlyIlePheCys 160
Db 2516 TGGAGGATATAGAGCTAAATTAACGAATGGCTTATTTTGTGTGTGTGTGTGTGTGTGT 2575
Qy 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleGly 180
Db 2576 GCCATTATAATGGAGGCTGCAACACAGCATTTGCAATAATATTTTCAAGATATATAGG 2635
Qy 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2636 GTTTTACAGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2695
Qy 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlnPheThrPhe 220
Db 2696 TTGTTTCTAGCCCTTGGAAATTTATTTCTTTTATATACATTTTTCCTTCAGGTTTCA 2755
Qy 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2756 GGCAGAGCTGGAGAGATCCTCACCAGCGCTCCGATACATGTTTCCGATCCATGCTC 2815
Qy 241 ArgGlnAspValSerTrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArg 260
Db 2816 AGACAGAGATGTGAGTTGGTTTGTGACCTAAACACCACTGGAGCATTCAGTACCAGG 2875
Qy 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
Db 2876 CTCGCCAATGATGCTGCTCAAGTTAAAGGGCTATAGTTCAGGCTTGTGTATATACC 2935
Qy 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
Db 2936 CAGATATAGCAATTTCTTGGGACAGGAATAATATATATCTTCTCATCTATGTTGCA 2995
Qy 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
Db 2996 ACACCTGTACTCTTAGCAATTTGACCATTCATTCGCAATAGCAGGATGTTTGAATGAA 3055
Qy 321 MetLeuSerGlyGlnAlaLeuLysAspLysGluLeuGluGlyAlaGlyLysIleAla 340
Db 3056 ATGTTGCTGGACAGCACTGAAAGATAAGAAAGAACTAGAAAGTCTGGGAAGATCGCT 3115
Qy 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGlu 360
Db 3116 ACTGAAGCAATAGAAACTTCCGAACCCGTTGTTTCTTTGACTCAGGAGCAGAAAGTTGA 3175
Qy 361 HisMetTyrAlaGlnSerLeuGln 368
Db 3176 CATATGATGCTCAGAGTTTGCGAG 3199
RESULT 9
US-08-752-447-1
Sequence 1, Application US/08752447
Patent No. 5994088
GENERAL INFORMATION:
APPLICANT: Mechetner, Eugene
APPLICANT: Roninson, Igor B
TITLE OF INVENTION: Methods and Reagents for Preparing and
TITLE OF INVENTION: Using Immunological Agents Specific for p-glycoprotein
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff Ltd.
STREET: 300 South Wacker Drive, Seventh Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible


```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,447
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5994088nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 95,1121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-9808
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..424
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 425..4264
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 4265...4669
; US-08-752-447-1
;
; Alignment Scores:
; Pred. No.: 3.95e-202 Length: 4669
; Score: 1644.00 Matches: 327
; Percent Similarity: 93.75% Conservative: 18
; Best Local Similarity: 88.86% Mismatches: 23
; Query Match: 0 Indels: 0
; DB: 2 Gaps: 0
;
; US-09-672-725c-6 (1-368) x US-08-752-447-1 (1-4669)
;
; QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValAlaLeuAspLysAlaArg 20
; DB ACCTTCAGCCCTGGACACAGAAAGCGAGCAGTGGTTCAGGTGGCTCTGGATAAGGCAGA 2155
;
; QY 21 LysGlyArgThrThrIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
; DB AAGGTCGGACCCACCATGTGTGATAGCTCATCTGTTTGTCTACAGTTCGTATGCTGACGTC 2215
;
; QY 41 IleAlaGlyPheAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
; DB ATCGCTGGTTTCGATGATGAGTCATTTGTGGAGAAAGGAATCATGATGAACATCATGAAA 2275
;
; QY 61 GluLysGlyIleThrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLeu 80
; DB GAGAAAGGCCATTTACTTCAAACTGTGCACAAATGCAGACAGCAAGAAATGAAGTTGAATTA 2335
;
; QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
; DB GAAATGCACCTGATGATCAAAAGTGAATGATGATGATGATGATGATGATGATGATGATGAT 2395
;
; QY 101 SerGlySerSerLeuLeuLysArgSerThrArgArgSerIleHisAlaProGlnGly 120
; DB TCAAGATCCAGTCTAATAAGAAAGAAAGTCAACTCGTAGGAGTGTCCGTGGATCACAAGCC 2455
;
; QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
; DB CAAACAGAAAGCTTAGTACCAAGAGGCTTGGATGAAAGTATACCTCCAGTTTCCTTT 2515
;
; QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTTPProTyrPheValValGlyIlePheCys 160
; DB TGGAGGATATGAAGCTAAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2575
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,447
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5994088nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 95,1121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-9808
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..424
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 425..4264
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 4265...4669
; US-08-752-447-1
;
; Alignment Scores:
; Pred. No.: 3.95e-202 Length: 4669
; Score: 1644.00 Matches: 327
; Percent Similarity: 93.75% Conservative: 18
; Best Local Similarity: 88.86% Mismatches: 23
; Query Match: 0 Indels: 0
; DB: 2 Gaps: 0
;
; US-09-672-725c-6 (1-368) x US-08-752-447-1 (1-4669)
;
; QY 161 AlaIleIleAsnGlyGlyLeuGlnProAlaPheSerIlePheSerArgIleIleGly 180
; DB GCCATTATAAATGAGCCCTGCAACAGCATTTGCAATATAATTTTCAAGATTATAGG 2635
;
; QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
; DB GTTTTACAAAGATTGATGATCCTGAAACAAACAGACAGAAATAGTAACCTGTTTTCACATA 2695
;
; QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
; DB TTTCTTCTAGCCCTTGGAAATTTATTTTATTTATTTTCTCTCAGGGTTTTCACATTT 2755
;
; QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeu 240
; DB GGCAGAGCTGGAGAGATCTCCACCAAGCGCTCCGATATGATGTTTTCGATCCATGCTC 2815
;
; QY 241 ArgGlnAspValSerTrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrArg 260
; DB AGACAGGATGTGAGTTGGTTTCATGACCTTAAACACACACCTGGAGCATTTGACTACCAG 2875
;
; QY 261 LeuAlaAsnAspAlaAlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThr 280
; DB CTGCCAATGATGCTGCTCAAGTTAAAGGGGCTATAGTTTCCAGGCTTCTGTAATTACC 2935
;
; QY 281 GlnAsnIleAlaAsnLeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeu 300
; DB CAGAAATATAGCAAACTTGGGACAGGAATAATATATATCTTCATCTATGTTGGCAACTA 2995
;
; QY 301 ThrLeuLeuLeuAlaIleValProIleIleAlaIleAlaGlyValValGluMetLys 320
; DB ACACCTGTTACTCTTAGCAATTTGACCATCATTTCAATAGCAGGAGTTGTTCAATGAAA 3055
;
; QY 321 MetLeuSerGlyClnAlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAla 340
; DB ATGTTTGGTGGACAGCACTGAAAGAAATAAGAAAGAACTAGAAAGGCTGCTGGGAAGATCGCT 3115
;
; QY 341 ThrGluAlaIleGluAsnPheArgThrValValSerLeuThrGlnGluLysPheGlu 360
; DB ACTGAAGCAATAGAAACCTCCGAACCGTTGTTCTTTGACTCAGACGAGCAAGTTTGA 3175
;
; RESULT 10
; US-09-120-513-1
; Sequence 1, Application US/09120513
; Patent No. 6025160
; GENERAL INFORMATION:
; APPLICANT: Brun, Kimberly
; APPLICANT: Chenery, Richard
; APPLICANT: Ellens, Harma
; APPLICANT: Field, John
; APPLICANT: Yue, Lin
; TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE
; TITLE OF INVENTION: SEQUENCES ENCODING RAT MDR1B2 AND
; TITLE OF INVENTION: SCREENING METHODS THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY:
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,513
; FILING DATE: 22-JUL-1998
```

```
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: King, William T
; REGISTRATION NUMBER: 30,954
; REFERENCE/DOCKET NUMBER: GP50008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5015
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4233 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-120-513-1

Alignment Scores:
Pred. No.: 5,11e-174 Length: 4233
Score: 1426.50 Matches: 275
Percent Similarity: 89.67% Conservative: 55
Best Local Similarity: 74.73% Mismatches: 37
Query Match: 77.70% Indels: 1
DB: 3 Gaps: 1

US-09-672-725c-6 (1-368) x US-09-120-513-1 (1-4233)

QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValAlaValAlaLeuAspLysAlaArg 20
Db 1691 ACCTGACGCTTGGACACAGAAAGCGGTGTTACGCGCGCTCTGGATAAGGCTAGA 1750

QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1751 GAAGCGCGGACCACTTGTATGATAGCTACCGCTTGTCTACAGTGGCAATGCTGACGTC 1810

QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1811 ATTGCTGTTTTGATGGTGTGTCATTTGTGGAGCAAGAAATCATGAAGAGCTCATGAAA 1870

QY 61 GluLysGlyIleThrPheLysLeuValThrMetGlnThrArgLysAsnGluIleAspLeu 80
Db 1871 GAGAAGGGCACTTTACTTCAAACTTGTCTACGACACAGCTAGAGGAAATGAAATGAACCA 1930

QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 1931 GGAATATATGCTTATGATCCCAAGATGACACTGGTGGCTCTGAGTGTACTTCAGAAAA 1990

QY 101 SerGlySerSerLeuLysArgSerThrArgSerThrArgSerThrIleHisAlaProGlnGly 120
Db 1991 TCAAAATCTCTTTAATA---AGGAGATCAATTCGCAAGATATCCACAGACAGCAAGAC 2047

QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
Db 2048 CAGAGAGAAGACTTAGTTCGAAAGGAGATGTGGATGAAGATGTGCTATGGTTTCCCTT 2107

QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProThrPheValValGlyIlePheCys 160
Db 2108 TGGCAGATCTTAAGCTAAATATAGTGAATGCCCTATTTAGTGTGGGTGACTTTGT 2167

QY 161 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 2168 GCTGTATAAATGGGTGCATACACACCAAGTGTGGTCCCATAGTGTGTTCAAAAGATGTAGGG 2227

QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2228 GTTTTTTCAAGAGACGACGACCATGAACCAACCAAGCAAGTGTACTTGTGTTTCCCTT 2287

QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 2288 CTCCTTCTGGTCATGGGAATGATTTCTTTGTTACGTACTTCTTCAAGGCTTCACATTT 2347

QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2751 GAAGCGCGGACCACTTGTGATAGCTACCGCTTGTCTTGTACTAGGAGCAGAACTTTGAA 2767

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: King, William T
; REGISTRATION NUMBER: 30,954
; REFERENCE/DOCKET NUMBER: GP50008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5015
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4233 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-120-513-1

Alignment Scores:
Pred. No.: 5,11e-174 Length: 4233
Score: 1426.50 Matches: 275
Percent Similarity: 89.67% Conservative: 55
Best Local Similarity: 74.73% Mismatches: 37
Query Match: 77.70% Indels: 1
DB: 3 Gaps: 1

US-09-672-725c-6 (1-368) x US-09-120-513-1 (1-4233)

QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValAlaValAlaLeuAspLysAlaArg 20
Db 1691 ACCTGACGCTTGGACACAGAAAGCGGTGTTACGCGCGCTCTGGATAAGGCTAGA 1750

QY 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1751 GAAGCGCGGACCACTTGTATGATAGCTACCGCTTGTCTACAGTGGCAATGCTGACGTC 1810

QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1811 ATTGCTGTTTTGATGGTGTGTCATTTGTGGAGCAAGAAATCATGAAGAGCTCATGAAA 1870

QY 61 GluLysGlyIleThrPheLysLeuValThrMetGlnThrArgLysAsnGluIleAspLeu 80
Db 1871 GAGAAGGGCACTTTACTTCAAACTTGTCTACGACACAGCTAGAGGAAATGAAATGAACCA 1930

QY 81 GluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAsp 100
Db 1931 GGAATATATGCTTATGATCCCAAGATGACACTGGTGGCTCTGAGTGTACTTCAGAAAA 1990

QY 101 SerGlySerSerLeuLysArgSerThrArgSerThrArgSerThrIleHisAlaProGlnGly 120
Db 1991 TCAAAATCTCTTTAATA---AGGAGATCAATTCGCAAGATATCCACAGACAGCAAGAC 2047

QY 121 GlnAspArgLysLeuGlyThrLysGluAspLeuAsnGluAsnValProProValSerPhe 140
Db 2048 CAGAGAGAAGACTTAGTTCGAAAGGAGATGTGGATGAAGATGTGCTATGGTTTCCCTT 2107

QY 141 TrpArgIleLeuLysLeuAsnSerThrGluTrpProThrPheValValGlyIlePheCys 160
Db 2108 TGGCAGATCTTAAGCTAAATATAGTGAATGCCCTATTTAGTGTGGGTGACTTTGT 2167

QY 161 AlaIleLeuAsnGlyGlyLeuGlnProAlaPheSerIleIlePheSerArgIleIleGly 180
Db 2168 GCTGTATAAATGGGTGCATACACCAAGTGTGGTCCCATAGTGTGTTCAAAAGATGTAGGG 2227

QY 181 IlePheThrArgAspGluAspProGluThrLysArgGlnIleSerAsnMetPheSerVal 200
Db 2228 GTTTTTTCAAGAGACGACGACCATGAACCAACCAAGCAAGTGTACTTGTGTTTCCCTT 2287

QY 201 LeuPheLeuValLeuGlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPhe 220
Db 2288 CTCCTTCTGGTCATGGGAATGATTTCTTTGTTACGTACTTCTTCAAGGCTTCACATTT 2347

QY 221 GlyLysAlaGlyGluIleLeuThrLysArgLeuArgLeuArgTyrMetValPheArgSerMetLeu 240
Db 2751 GAAGCGCGGACCACTTGTGATAGCTACCGCTTGTCTTGTACTAGGAGCAGAACTTTGAA 2767
```



```
Db 2050 GACGCGCTTGGAGGACCGCGACGCTGAGGATCTCACGAATGCAGATATTCGCAAAATCAA 2109
Qy 98 ProLysAspSerGlySerSer-----LeuIleLysArgArgSer 110
Db 2110 ACTGCGTCAAGCGCATCTCCGATCCTCGAGGAGAAACCCACAAACCATTCACCGCAGGGCC 2169
Qy 111 ThrArgArgSerIleHisAlaProGlnGlnGlnAspArgLysLeuGlyThrLysGluAsp 130
Db 2170 ACCCAAGTCTGTTCCAGCGGATTCCTTAAAGACCC----- 2211
Qy 131 LeuAsnGluAsnValProProValSerPheThrArgIleLeuLys-----Leu 146
Db 2212 ---CCGCAAACTCCGAATACTATTATGACGCTGCTCAAAATTTGTTGCTTCCTTC 2268
Qy 147 AsnSerThrGluTrpProTyrPheValGlyIlePheCysAlaIleAsnGlyGly 166
Db 2269 AACGGCCTGAATVCCGATACGCTATCGTCTTCTTCAGTGTAGTGTGTGT 2328
Qy 167 LeuGlnProAlaPheSerIlePheSerArgIleIleGlyPheThrArgAspGlu 186
Db 2329 GGCAACCCACGACGAGTCTATATGCTAAAGCCATCAGCACACTCTCGCTCCAGAA 2388
Qy 187 AspProGluThrLys---ArgGlnIleSerAsnMetPheSerValLeuPheLeuValLeu 205
Db 2389 TCACATATAGCAAGCTTCGACATGATCGGATTTCTGTCATGTATGTTTCGTGGTT 2448
Qy 206 GlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLysAlaGlyGlu 225
Db 2449 GGTATCATTCAGTTTATCAGCGAGTCAACCAATGGTGTGCTCATTTGCCGTATGCTCCGAG 2508
Qy 226 IleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGlnAspValSer 245
Db 2509 AGACTTATTCGTCGGGAGAGACACTGCTTTCGGAGCATCTCCGTCGAAGACATTTGCT 2568
Qy 246 TrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArgLeuAlaAsnAspAla 265
Db 2569 TTCTTTTCAAGAGAGAGATAGACACGGCGCTGTGACCTTCTTCCTGTCACCGGAGAGC 2628
Qy 266 AlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsn 285
Db 2629 AAGCATCTCTCCGGTGTAGCGGTGTGACCTAGGCACGATCTTGATGACCTCCACGACC 2688
Qy 286 LeuGlyThrGlyIleIleIleSerLeuIleThrGlyTrpGlnLeuThrLeuLeuLeu 305
Db 2689 CTAGAGCGGTATCATATTATTCGCTCGGATTTGGTGGTAAATTTGGCTTTGTTGATC 2748
Qy 306 AlaIleValProIleAlaIleAlaGlyValValGluMetLysMetLeuSerGlyGln 325
Db 2749 TCGGTTGGCGGTTCTCTGGCATGCGGTTCTTACCGATCTATATGCTAGCCAGTTT 2808
Qy 326 AlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGlu 345
Db 2809 CAATCAGCGCTCAAGCTGCTTATAGGGATCTGCAAACTTTGCTTCGAGGCTACATCG 2868
Qy 346 AsnPheArgThrValSerLeuThrGlnGluGlnLysPheGluHisMetTyrAlaGln 365
Db 2869 TCTATCCGACAGTTCGCTTCAATACCCGGGAAGGATGCTGTGGGAGATTTTACCATGCC 2928
Qy 366 SerLeu 367
Db 2929 CAGCTT 2934
```

RESULT 13
US-08-996-545-3
; Sequence 3, Application US/08996545
; Patent No. 5928898
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene atrd of
; TITLE OF INVENTION: Aspergillus nidulans

NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: U.S.
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/996,545
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Webster, Thomas D.
REGISTRATION NUMBER: 39,872
REFERENCE/DOCKET NUMBER: X-11766
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3334
TELEFAX: 317-276-2763
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4002 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-996-545-3

Alignment Scores:
Pred. No.: 8,55e-66 Length: 4002
Score: 591.00 Matches: 141
Percent Similarity: 53.93% Conservative: 65
Best Local Similarity: 36.91% Mismatches: 154
Query Match: 32.19% Indels: 22
DB: 2 Gaps: 6

US-09-672-725C-6 (1-368) x US-08-996-545-3 (1-4002)
Qy 1 ThrSerAlaLeuAspThrGluSerGluAlaValGlnValAlaLeuAspLysAlaArg 20
Db 1810 ACUUCGGCCUUGGACACAAAUCGAAAGCGUGGUCAAGCAGCUUUGAGAGGCGCAGCU 1869
Qy 21 LysGlyArgThrThrIleValIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1870 GAAGCGCGAAGCUUUAUUGUGAUGCUCAUCGCCUUCACGAUCAAACGCGCAGCAAC 1929
Qy 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1930 AUUGUGGUUCUGUCAAAGGAAAAUUGUGAACAAGAACUCACGAUAAUUGGUUGAC 1989
Qy 61 GluLysGlyIleTyrPheLysLeuValThrMetGln-----ThrArgLysAsnGluIle 78
Db 1990 CGCGAGGCGCUUUAUCGCAAAUCUUGUGAGGCGUACGUCAUUAUAGAACAGAGGAAGCU 2049
Qy 79 Asp---LeuGluAsnAlaThrGlyGluSerLysSerGluSerAlaLeuGluMetSer 97
Db 2050 GAGCCUUGGAGGAGCGCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2109
Qy 98 ProLysAspSerGlySerSer-----LeuIleLysArgArgSer 110
Db 2110 ACUCGCUCAAGCGCAUCAUCCGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2169
Qy 111 ThrArgArgSerIleHisAlaProGlnGlnGlnAspArgLysLeuGlyThrLysGluAsp 130
Db 2170 ACCCAAGGUCUGUUCGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2211

```
QY 131 LeuAsnGluAsnValProProValSerPheTrpArgIleLeuLys-----Leu 146
Db 2212 ---CCGAAACAAACUCCGAAAUACUUAUUGAGCGGCGUCAAUUUGUUGUCCUUC 2268
QY 147 AsnSerThrGluTrpProTyrPheValValGlyIlePheCysAlaIleIleAsnGlyGly 166
Db 2269 AACCGCCGGAUACUCCGAAUACUUGAGCGGCGUUAUUGUUGUUGUUGUUGUUGU 2328
QY 167 LeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePheThrArgAspGlu 186
Db 2329 GCCCAACCCAGCAGAGUUAUUGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2388
QY 187 AspProGluThrLys---ArgGlnIleSerAsnMetPheSerValLeuPheLeuValIle 205
Db 2389 UCACAAUAGCAAGCUUCGACAUUGAGUUGUUGUUGUUGUUGUUGUUGUUGUUGU 2448
QY 206 GlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLysAlaGlyGlu 225
Db 2449 GGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 2508
QY 226 IleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGlnAspValSer 245
Db 2509 AGACUUAUUGCGCGGAGAGACACUCCUUGCGAGGAGUUCUCCGACAGACUUGCU 2568
QY 246 TrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArgLeuAlaAsnAspAla 265
Db 2569 UUCUUGACAAGGAGAGAAUAGCAGCGGCGUUGUUGUUGUUGUUGUUGUUGUUGU 2628
QY 266 AlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsn 285
Db 2629 AAGCAUCUUCGCGGUGUUGAGCGGUGACUACGACGACGACGACGACGACGACGAC 2688
QY 286 LeuGlyThrGlyIleIleIleSerLeuIleTyrGlnLeuThrGlnLeuThrLeuLeu 305
Db 2689 CUAGAGCGGCUUAUUAUUGCGGCGUUGGCGGAGUUGGCGGAGUUGGCGGAGUUG 2748
QY 306 AlaIleValProfileIleAlaIleAlaIleAlaGlyValValGluMetLysMetLeuSerGlyGln 325
Db 2749 UCGGUGUGCGGUGUUGUUGGUGGUGUUGUUGUUGUUGUUGUUGUUGUUGUUGU 2808
QY 326 AlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGlu 345
Db 2809 CAUUCAGCGUCCAGCUUGCUUAUUGAGGAGUUGCUACAAACUUAUUGCUUGCG 2868
QY 346 AsnPheArgThrValSerLeuThrGlnGlnLysPheGluHisMetTyrAlaGln 365
Db 2869 UCUAUCGCGAGUUGGUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2928
QY 366 SerLeu 367
Db 2929 CAGCUU 2934

RESULT 14
US-09-328-320-1
; Sequence 1, Application US/09328320
; Patent No. 6228615
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul L.
; APPLICANT: de Waard, Maarten A.
; APPLICANT: Peery, Robert B.
; APPLICANT: Andrade, Alan C.
; TITLE OF INVENTION: Multiple Drug Resistance Gene atrd of
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/328,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/996,545
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11766
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-2763
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4002
; US-09-328-320-1

Alignment Scores:
Pred. No.: 8,55e-66 Length: 4002
Score: 591.00 Matches: 141
Percent Similarity: 53.93% Conservative: 65
Best Local Similarity: 36.91% Mismatches: 154
Query Match: 32.19% Indels: 22
DB: 4 Gaps: 6

US-09-672-725C-6 (1-369) x US-09-328-320-1 (1-4002)
QY 1 ThrSerAlaLeuAspThrGluSerGluAlaValValGlnValAlaLeuAspLysAlaArg 20
Db 1810 ACTTCGGCCTTGGACACAAATCCGAAGCGTGGTTCAAGCAGCTTGGAGAGGCGAGCT 1869
QY 21 LysGlyArgThrThrIleAlaHisArgLeuSerThrValArgAsnAlaAspVal 40
Db 1870 GAAGCGCGAACTACTATTGTGATCGCTCATCGCTTTCACGATCAAAACGCGCACAC 1929
QY 41 IleAlaGlyPheAspAspGlyValIleValGluLysGlyAsnHisAspGluLeuMetLys 60
Db 1930 ATTGTGGTTCGTCAATGGCAAAATGCTGAACAGGAACTCAGCATGAATGGTTGAC 1989
QY 61 GluLysGlyIleTyrPheLysLeuValThrMetGln-----ThrArgGlyAsnGluIle 78
Db 1990 CGCGGAGGCGCTTATCGCAACTTGTGGAGGCTCAACGATCAATCAACAGAGGAAGCT 2049
QY 79 Asp---LeuGluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSer 97
Db 2050 GACGCGCTTGGAGGAGCGCGACGCTGAGGATCTCAGGAATGCAGATATTGCCAAATCAA 2109
QY 98 ProLysAspSerGlySerSer-----LeuIleLysArgArgSer 110
Db 2110 ACTGCGCTAAGCGCATCATCCGATCTCGACGGAACCCACCAACCATCCGCGACGGGC 2169
QY 111 ThrArgArgSerIleHisAlaProGlnGlyGlnAspArgLysLeuGlyThrLysGluAsp 130
Db 2170 ACCCAAGCTGTTCAGCGCGGATCTTCTTAAAGACCC-----2211
QY 131 LeuAsnGluAsnValProProValSerPheTrpArgIleLeuLys-----Leu 146
Db 2212 ---CCGAAACAAACUCCGAAAUACUUAUUGAGCGGCGUCAAUUUGUUGUCCUUC 2268
```

Qy	147	AsnSerThrGluTrpProTyrPheValValGlyLeuPheCysAlaIleIleLeuAsnGlyGly	166
Db	2269	: : : :	
Qy	167	LeuGlnProAlaPheSerIleIlePheSerArgIleIleGlyIlePheThrArgAspGlu	186
Db	2329	: : : : : : : :	
Qy	187	AspProGluThrLys---ArgGlnIleSerAsnMetPheSerValLeuPheLeuValLeu	205
Db	2389	: : : : : : : : : :	
Qy	206	GlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLysAlaGlyGlu	225
Db	2449	GGTATCATTCAGCTTTTACACGAGTCAACCAATGGTGCTGCATTGCCGTATGCTCCGAG	2508
Qy	226	IleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGlnAspValSer	245
Db	2509	: : : : : : : :	
Qy	246	TripPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArgLeuAlaAsnAspAla	265
Db	2569	: : : : : : : : : :	
Qy	266	AlaGlnValLysAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsn	285
Db	2629	AAGCATCTCTCCCGGTGTAGCGGTGTACTCTAGCAGCATCTTGATGACCTCCACGACC	2688
Qy	286	LeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeu	305
Db	2689	CTAGGAGCGGTATCATTTATTCCTCGCGATTGGGTGGAATTTGGCTTAGTTTGTATC	2748
Qy	306	AlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMetLeuSerGlyGln	325
Db	2749	: : : : : : : : : :	
Qy	326	AlaLeuAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGlu	345
Db	2809	: : : : : : : : : :	
Qy	346	AsnPheArgThrValValSerLeuThrGlnGluGlnLysPheGluHisMetTyrAlaGln	365
Db	2869	TCATTCGCGACATTCGGTTCATTAACCGGAAAGGGATGTCTGGGAGATTTACCATGCC	2928
Qy	366	SerLeu 367	
Db	2929	CAGCTT 2934	
RESULT 15			
US-09-328-320-3			
; Sequence 3, Application US/09328320			
; Patent No. 6228615			
; GENERAL INFORMATION:			
; APPLICANT: Skatrud, Paul L.			
; APPLICANT: de Waard, Maarten A.			
; APPLICANT: Peery, Robert B.			
; APPLICANT: Andrade, Alan C.			
; TITLE OF INVENTION: Multiple Drug Resistance Gene atRD of			
; TITLE OF INVENTION: Aspergillus nidulans			
; NUMBER OF SEQUENCES: 3			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Eli Lilly and Company			
; STREET: Lilly Corporate Center			
; CITY: Indianapolis			
; STATE: Indiana			
; COUNTRY: U.S.			
; ZIP: 46285			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			

```

Qy 187 AspProGluThrLys---ArgGlnIleSerAsnMetPheSerValLeuPheLeuValLeu 205
Db 2389 UCACAAUAUAGCAGCUUCGACAUUGCGGAUUCUGGUCAUUGAUUGUUCUGUGGU 2448
Qy 206 GlyIleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLysAlaGlyGlu 225
Db 2449 GGAUCAUUCAGUUUUCACGACGACACCAACAAUGGUGCAUUGCCGUAUGCCGAG 2508
Qy 226 IleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGlnAspValSer 245
Db 2509 AGACUUAUUCGCGCGGAGAGACACGCUUCUGGAGCAUACUCCGCAAGACAUUGCU 2568
Qy 246 TrpPheAspLeuLysAsnThrThrGlyAlaLeuThrThrArgLeuAlaAsnAspAla 265
Db 2569 UUCUUUGACAGGAGAGAAAGACCGGCGCUCUGACCCUUCUUCUGUCCACCAGAGACG 2628
Qy 266 AlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsn 285
Db 2629 AAGCAUCUCGCGGUUAGCGUGUGACUCUAGGCACGAUCUUGAGCCUCCACCAGACC 2688
Qy 286 LeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeu 305
Db 2689 CUAGGAGCGGCUAUCAUUAUUGCCUGGCGGAUUGGGUGGAAUUGGCCUUAUGUAUC 2748
Qy 306 AlaIleValProIleIleAlaIleAlaGlyValValGluMetLysMetLeuSerGlyGln 325
Db 2749 UCGGUUGUGCGGUUCUCCGCAUGCGGUUUCUACCGAUUCUAUUGCUAGCCCGGUU 2808
Qy 326 AlaLeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGlu 345
Db 2809 CAAUCACCCUCCAGCUUGCUUAUGAGGGAUCUGCAACUUGCUUGCGGAGCUACAUCG 2868
Qy 346 AsnPheArgThrValValSerLeuThrGlnGlnLysPheGluHisMetTyrAlaGln 365
Db 2869 UCUAUCCGCACAGUUGCGUCAUAUAAACCGGGAAGGAUGUCUGGAGAUUUUACCAUGCC 2928
Qy 366 SerLeu 367
Db 2929 CAGCUU 2934

```

Search completed: November 6, 2002, 19:17:12
Job time : 68.8217 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:33:09 ; Search time 5.76361 Seconds
(without alignments)
6135.192 Million cell updates/sec

Title: US-09-672-725C-6
Perfect score: 1836
Sequence: 1 TSLDTESEAVVQVALDKAR.....TVVSLTQEQKFEHMYAQSLQ 368

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pirl.*
2: pirl2.*
3: pirl3.*
4: pirl4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1672	91.1	1280	1 DVHU1	multidrug resistan
2	1589	86.5	1276	1 DVHV1C	multidrug resistan
3	1584	86.3	1104	1 DVMS1A	multidrug resistan
4	1584	86.3	1276	2 A34786	multidrug resistan
5	1418.5	77.3	1277	2 JH0502	p-glycoprotein - r
6	1401.5	76.3	1276	1 DVMS1	multidrug resistan
7	1248.5	68.0	1279	1 DVHU3	multidrug resistan
8	1220.5	66.5	1276	1 DVMS2	multidrug resistan
9	1207.5	65.8	1278	2 S41646	multidrug resistan
10	1204.5	65.6	1281	2 I48123	p-glycoprotein iso
11	1148	62.5	1287	2 S55692	multidrug resistan
12	1071.5	58.4	655	1 DVHV2C	multidrug resistan
13	782.5	42.6	1321	2 T42228	p-glycoprotein sis
14	782.5	42.6	1321	2 T42842	multidrug resistan
15	677.5	36.9	1275	2 T31073	bile salt transpor
16	666	36.3	1283	2 A47377	multidrug resistan
17	653.5	35.6	1294	2 T19982	multidrug resistan
18	632.5	34.4	1289	2 D87789	hypothetical prote
19	632	34.4	1292	2 T48007	protein C34G6.4 [i
20	613	33.4	1278	2 E86155	p-glycoprotein hom
21	612.5	33.4	1286	2 T02187	probable ABC trans
22	605.5	33.0	1302	2 A41249	probable ABC trans
23	598.5	32.6	1302	2 B41249	multidrug resistan
24	595.5	32.4	1230	2 E85023	multidrug resistan
25	593	32.3	1229	2 F86155	probable p-glycopr
26	589.5	32.1	1321	2 T23476	probable ABC trans
27	589.5	32.1	1321	2 S27337	hypothetical prote
28	575.5	31.3	1229	2 D85023	multidrug resistan
29	575.5	31.3	1229	2 T52319	p-glycoprotein-lik

30	558	30.4	1286	2 A42150	p-glycoprotein ppp
31	554.5	30.2	1266	2 T22090	hypothetical prote
32	551.5	30.0	1254	2 T30855	multidrug resistan
33	544.5	29.7	1222	2 T14805	hypothetical prote
34	544.5	29.7	1307	2 T30882	multidrug resistan
35	544.5	29.7	1323	2 H85202	hypothetical prote
36	543	29.6	1268	2 T22094	hypothetical prote
37	524	28.5	1233	2 T04251	p-glycoprotein 2 -
38	524	28.5	1245	2 G86404	probable p-glycopr
39	521.5	28.4	1408	2 T43261	multidrug resistan
40	504	27.5	1232	2 T06165	multidrug resistan
41	503	27.4	1407	1 T00558	probable ABC trans
42	502	27.3	1310	2 S30328	multidrug resistan
43	501.5	27.3	1247	2 F86405	probable p-glycopr
44	501	27.3	1302	2 S30327	multidrug resistan
45	495.5	27.0	1408	2 T47671	p-glycoprotein-lik

ALIGNMENTS

RESULT 1

DVHU1

multidrug resistance protein 1 - human

N:Alternate names: p-glycoprotein 1

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1990 #sequence_revision 18-Aug-1995 #text_change 19-Jan-2001

C:Accession: A34914; PS0162; S15500; A25059; S43838; I52238; I65204

R:Chen, C.; Clark, D.; Ueda, K.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.

J. Biol. Chem. 265, 506-514, 1990

A>Title: Genomic organization of the human multidrug resistance (MDR1) gene and origi

A:Reference number: A34914; MUID:90094448

A:Accession: A34914

A:Molecule type: DNA

A:Residues: 1-1280 <CHE>

A:Cross-references: GB:M29447; GB:J05168; NID:g187496; PIDN:AAA59576.1; PID:g386862

R:Kioka, N.; Yamano, Y.; Komano, T.; Ueda, K.

submitted to JIPID, April 1991

A:Reference number: PS0162

A:Accession: PS0162

A:Molecule type: DNA

A:Residues: 1-22 <KIO>

R:Kioka, N.; Yamano, Y.; Komano, T.; Ueda, K.

submitted to the EMBL Data Library, April 1991

A:Description: Transcriptional regulation of multidrug resistance gene (MDR1) express

A:Reference number: S15500

A:Accession: S15500

A:Molecule type: DNA

A:Residues: 1-22, 'R' <K12>

A:Cross-references: EMBL:X58723; NID:g34522; PIDN:CAA41558.1; PID:g34523

R:Chen, C.; Chin, J.E.; Ueda, K.; Clark, D.P.; Pastan, I.; Gottesman, M.M.; Roninson,

Cell 47, 381-389, 1986

A>Title: Internal duplication and homology with bacterial transport proteins in the m

A:Reference number: A25059; MUID:87028230

A:Accession: A25059

A:Molecule type: mRNA

A:Residues: 1-184, 'V', 186-1280 <CH2>

A:Cross-references: GB:ML4758; NID:g187468; PIDN:AAA59575.1; PID:g307180

R:Chambers, T.C.; Pohl, J.; Glass, D.B.; Kuo, J.F.

Biochem. J. 299, 309-315, 1994

A>Title: Phosphorylation by protein kinase C and cyclic AMP-dependent protein kinase

A:Reference number: S43838; MUID:94220047

A:Accession: S43838

A:Molecule type: protein

A:Residues: 656-689 <CHA>

R:Geckeler, V.; Weger, S.; Probst, H.

Biochem. Biophys. Res. Commun. 169, 796-802, 1990

A>Title: mdrl/p-glycoprotein gene segments analyzed from various human leukemic cell

A:Reference number: 152238; MUID:90290529

A:Accession: 152238

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 178-215 <RES>

A:Cross-references: GB:M37724; NID:g183537; PIDN:AAA8047.1; PID:g553314
A:Accession: 165204
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 800-856 <RE2>
A:Cross-references: GB:M37725; NID:g183538; PIDN:AAA8048.1; PID:g553315
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: GDB:PGY1; MDR1
A:Cross-references: GDB:120712; OMIM:171050
A:Map position: 7q21-7q21
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; phosphoprotein;
F:1-638,653-1280/Region: duplication
F:49-350/Domain: hydrophobic <HB1>
F:351-637/Domain: hydrophilic <HL1>
F:410-604/Domain: ATP-binding cassette homology <ABC1>
F:427-434/Region: nucleotide-binding motif A (P-loop)
F:551-555/Region: nucleotide-binding motif B
F:638-708/Domain: linker <LIN>
F:709-993/Domain: linker <LIN>
F:994-1280/Domain: hydrophobic <HB2>
F:1053-1249/Domain: ATP-binding cassette homology <ABC2>
F:1070-1077/Region: nucleotide-binding motif A (P-loop)
F:1196-1200/Region: nucleotide-binding motif B
F:91-94-99/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:433/Binding site: ATP (Lys) #status predicted
F:661,667,671/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status exp
F:667,671/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status
F:1076/Binding site: ATP (Lys) #status predicted

Query Match 91.1%; Score 1672; DB 1; Length 1280;
Best Local Similarity 90.5%; Pred. No. 2.2e-126;
Matches 333; Conservative 16; Mismatches 19; Indels 0; Gaps 0;

Qy 1 TSALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 60
Db 558 TSALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 617
Qy 61 EKGIFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRSTRSHAPQG 120
Db 618 EKGIFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRSTRSHAPQG 677
Qy 121 QDRKLGTEDLNENPVVPSFWRILKNSTWPFYVVGIFCAIINGGLQPAFISIRIG 180
Db 678 QDRKLSTKEALDESIPPVPSFWRILKNSTWPFYVVGIFCAIINGGLQPAFISIRIG 737
Qy 181 IFTREDEPETKRLQISNMFSLFLVGLIISITFFLQGFPGKAGEILTTLKRLYVWFRSML 240
Db 738 VFTRIDDPETKRLQISNMFSLFLVGLIISITFFLQGFPGKAGEILTTLKRLYVWFRSML 797
Qy 241 RODYSWFDLKNNTGALTTLRLANDAAQVKAIGSLAVITONTANLGTGIIISLIYQWL 300
Db 798 RODYSWFDLKNNTGALTTLRLANDAAQVKAIGSLAVITONTANLGTGIIISLIYQWL 857
Qy 301 TLLLLAIVPIIATAGVYEMKLSQALKKKLEGGAGKIATEIENFRVVSITQEQKE 360
Db 858 TLLLLAIVPIIATAGVYEMKLSQALKKKLEGGAGKIATEIENFRVVSITQEQKE 917
Qy 361 HMYAQSLQ 368
Db 918 HMYAQSLQ 925

RESULT 2
DVHVIC
multidrug resistance protein 1 - Chinese hamster
N:Alternate names: P-glycoprotein pgp1
C:Species: Cricetus griseus (Chinese hamster)
C:Date: 31-Dec-1990 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A38696; C38696; A27126; S33768; I52823
S:Devina, S.E.; Hussain, A.; Davide, J.P.; Melera, P.W.

J. Biol. Chem. 266, 4545-4555, 1991
A:Title: Full length and alternatively spliced pgp-1 transcripts in multidrug-resista
A:Reference number: A38696; MUID:91154265
A:Accession: A38696
A:Molecule type: mRNA
A:Residues: 1-1276 <DEV>
A:Cross-references: GB:M59253; NID:g191154; PIDN:AAA37004.1; PID:g191155
A:Accession: C38696
A:Molecule type: mRNA
A:Residues: 108-1276
A:Cross-references: GB:M59254; NID:g191156; PIDN:AAA37005.1; PID:g191157
A:Experimental source: clone ADX185
A:Accession: B38696
A:Molecule type: mRNA
A:Residues: 1-32,771-1276 <DE2>
A:Cross-references: GB:M59252; NID:g191152; PIDN:AAA37003.1; PID:g191153
A:Experimental source: clone ADX124
R:Endicott, J.A.; Juranka, P.F.; Sarangi, F.; Gerlach, J.H.; Deuchars, K.L.; Ling, V.
Mol. Cell. Biol. 7, 4075-4081, 1987
A:Title: Simultaneous expression of two P-glycoprotein genes in drug-sensitive Chines
A:Reference number: A27126; MUID:88122132
A:Accession: A27126
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 706-1276 <END>
A:Cross-references: GB:M17897; NID:g191158; PIDN:AAA37006.1; PID:g191159
R:Zastawny, R.L.; Ling, V.
Biochim. Biophys. Acta 1173, 303-313, 1993
A:Title: Structural and functional analysis of 5' flanking and intron 1 sequences of
A:Reference number: S33768; MUID:93305724
A:Accession: S33768
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-21 <ZAS>
A:Cross-references: EMBL:L03286
R:Teeter, L.D.; Eckersberg, T.; Tsai, Y.; Kuo, M.T.
Cell Growth Differ. 2, 429-437, 1991
A:Title: Analysis of the Chinese hamster P-glycoprotein/multidrug resistance gene pgp
A:Reference number: I52823; MUID:92088970
A:Accession: I52823
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-21 <RES>
A:Cross-references: GB:S81975; NID:g240862
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c
structurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: pgp1
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: alternative splicing; ATP; duplication; glycoprotein; nucleotide binding;
F:407-601/Domain: ATP-binding cassette homology <ABC1>
F:424-431/Region: nucleotide-binding motif A (P-loop)
F:548-552/Region: nucleotide-binding motif B
F:1050-1246/Domain: ATP-binding cassette homology <ABC2>
F:1067-1074/Region: nucleotide-binding motif A (P-loop)
F:1193-1197/Region: nucleotide-binding motif B
F:87,91,96/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:430/Binding site: ATP (Lys) #status predicted
F:1073/Binding site: ATP (Lys) #status predicted

Query Match 86.5%; Score 1589; DB 1; Length 1276;
Best Local Similarity 85.1%; Pred. No. 1.1e-119;
Matches 313; Conservative 29; Mismatches 26; Indels 0; Gaps 0;

Qy 1 TSALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 60
Db 555 TSALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 614
Qy 61 EKGIFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRSTRSHAPQG 120
Db 615 EKGIFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRSTRSHAPQG 674
Qy 121 QDRKLGTEDLNENPVVPSFWRILKNSTWPFYVVGIFCAIINGGLQPAFISIRIG 180

Db 675 QDRKLTKEALDEVPPISEFWIRKLNSSEWPFVVGIFCAIVNGALQPAFSIIFSKVVG 734
QY 181 IFTREDDEPTKROIISNMFSLVFLVGLIISFTIFFLQGTFFGKAGELTKRLRYMVFRSML 240
Db 735 VFTRNTDDETKKHSNLSLFLIUGVISFTIFFLQGTFFGKAGELTKRLRYMVFRSML 794
QY 241 RQDVSWFDDLNKTTGALTTRLANDAAQVKGAIGSRSLAVITONIANLGTGIIISLIYGWOL 300
Db 795 RQDVSWFDDNPKNTTGALTTRLANDAGQVGATGARLAVITONIANLGTGIIISLIYGWOL 854
QY 301 TLLLLAIVPIIATAGVVEKMLSGQALKDKKKELEGAGKIATEAIEFNFTVVSLSLTOEQKFE 360
Db 855 TLLLLAIVPIIATAGVVEKMLSGQALKDKKKELEGAGKIATEAIEFNFTVVSLSLTOEQKFE 914
QY 361 HMYAQSLQ 368
Db 915 NMIAQSLQ 922
RESULT 3
DVMSIA
multidrug resistance protein 1a - mouse (fragment)
N:Alternate names: P-glycoprotein 1a
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: A34175
J: Hsu, S.I.H.; Lothstein, L.; Horwitz, S.B.
J: Biol. Chem. 264, 12053-12062, 1989.
A:Title: Differential overexpression of three mdr gene family members in multidrug-resistant
A:Reference number: A34175; MUID:89308614
A:Accession: A34175
A:Molecule type: mRNA
A:Residues: 1-1104 <HSD>
A:Cross-references: GB:M24417; GB:J04839; NID:g200329; PIDN:AAA03243.1; PID:g200330
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic antitumor drugs.
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane pr
F:234-428/Domain: ATP-binding cassette homology <ABC1>
F:251-258/Region: nucleotide-binding motif A (P-loop)
F:375-379/Region: nucleotide-binding motif B
F:877-1073/Domain: ATP-binding cassette homology <ABC2>
F:894-901/Region: nucleotide-binding motif A (P-loop)
F:1020-1024/Region: nucleotide-binding motif B
F:257/Binding site: ATP (Lys) #status predicted
F:900/Binding site: ATP (Lys) #status predicted
Query Match 86.3%; Score 1584; DB 1; Length 1104;
Best Local Similarity 86.1%; Pred. No. 2.2e-119;
Matches 317; Conservative 22; Mismatches 29; Indels 0; Gaps 0;
QY 1 TSALDTESEAVVQVALDKARKGRITIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 60
Db 382 TSALDTESEAVVQALDKAREGRITIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMR 441
QY 61 EKGIFYKLVTMOTRGNEDLNATGESKSDALEMSPKDSGLIKRRSPRRSIHAPOG 120
Db 442 EKGIFYKLVTMOTRGNEDLNATGESKSDALEMSPKDSGLIKRRSPRRSIHAPOG 501
QY 121 QDRKLTGKEDLNENPVPSFWRILKLNSTWPFVVGIFCAIINGGLQPAFSIIFSRIIG 180
Db 502 QDRKLTKEALDEVPVPAFWRILKLNSTWPFVVGIFCAIINGGLQPAFSIIVSKVVG 561
QY 181 IFTREDDEPTKROIISNMFSLVFLVGLIISFTIFFLQGTFFGKAGELTKRLRYMVFRSML 240
Db 562 VFTNGGPPETQRQNSLFLILIGIISFTIFFLQGTFFGKAGELTKRLRYMVFRSML 621
QY 241 RQDVSWFDDLNKTTGALTTRLANDAAQVKGAIGSRSLAVITONIANLGTGIIISLIYGWOL 300
Db 622 RQDVSWFDDNPKNTTGALTTRLANDAAQVKGATGARLAVITONIANLGTGIIISLIYGWOL 681
QY 301 TLLLLAIVPIIATAGVVEKMLSGQALKDKKKELEGAGKIATEAIEFNFTVVSLSLTOEQKFE 360

Db 682 TLLLLAIVPIIATAGVVEKMLSGQALKDKKKELEGAGKIATEAIEFNFTVVSLSLTOEQKFE 741
QY 361 HMYAQSLQ 368
Db 742 TMYAQSLQ 749
RESULT 4
A34786
multidrug resistance protein 1a - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jul-1990 #sequence_revision 13-Jul-1990 #text_change 02-Feb-2001
C:Accession: A34786; A35671
R:Devault, A.; Gros, P.
Mol. Cell. Biol. 10, 1652-1663, 1990
A:Title: Two members of the mouse mdr gene family confer multidrug resistance with ov
A:Reference number: A34786; MUID:90205845
A:Accession: A34786
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1276 <DEV>
A:Cross-references: GB:M30697; NID:gl99111; PIDN:AAA39517.1; PID:g387429
R:Hsu, S.I.H.; Cohen, D.; Kirschner, L.S.; Lothstein, L.; Hartstein, M.; Horwitz, S.B.
Mol. Cell. Biol. 10, 3596-3606, 1990
A:Title: Structural analysis of the mouse mdrla (P-glycoprotein). promoter reveals the
A:Reference number: A35671; MUID:90287150
A:Accession: A35671
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-525, 'HV', 528-938, 'S', 940-1035, 'V', 1037-1276 <HSU>
A:Cross-references: GB:M33581; NID:gl99104; PIDN:AAA39514.1; PID:g387427
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; membrane protein; nucleotide binding; P-1
F:406-600/Domain: ATP-binding cassette homology <ABC1>
F:423-431/Region: nucleotide-binding motif A (P-loop)
F:547-551/Region: nucleotide-binding motif B
F:1049-1245/Domain: ATP-binding cassette homology <ABC2>
F:1066-1074/Region: nucleotide-binding motif A (P-loop)
F:1192-1196/Region: nucleotide-binding motif B
F:429/Binding site: ATP (Lys) #status predicted
F:1072/Binding site: ATP (Lys) #status predicted
Query Match 86.3%; Score 1584; DB 2; Length 1276;
Best Local Similarity 86.1%; Pred. No. 2.7e-119;
Matches 317; Conservative 22; Mismatches 29; Indels 0; Gaps 0;
QY 1 TSALDTESEAVVQVALDKARKGRITIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 60
Db 554 TSALDTESEAVVQALDKAREGRITIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMR 613
QY 61 EKGIFYKLVTMOTRGNEDLNATGESKSDALEMSPKDSGLIKRRSPRRSIHAPOG 120
Db 614 EKGIFYKLVTMOTRGNEDLNATGESKSDALEMSPKDSGLIKRRSPRRSIHAPOG 673
QY 121 QDRKLTGKEDLNENPVPSFWRILKLNSTWPFVVGIFCAIINGGLQPAFSIIFSRIIG 180
Db 674 QDRKLTKEALDEVPVPAFWRILKLNSTWPFVVGIFCAIINGGLQPAFSIIVSKVVG 733
QY 181 IFTREDDEPTKROIISNMFSLVFLVGLIISFTIFFLQGTFFGKAGELTKRLRYMVFRSML 240
Db 734 VFTNGGPPETQRQNSLFLILIGIISFTIFFLQGTFFGKAGELTKRLRYMVFRSML 793
QY 241 RQDVSWFDDLNKTTGALTTRLANDAAQVKGAIGSRSLAVITONIANLGTGIIISLIYGWOL 300
Db 794 RQDVSWFDDNPKNTTGALTTRLANDAAQVKGATGARLAVITONIANLGTGIIISLIYGWOL 853
QY 301 TLLLLAIVPIIATAGVVEKMLSGQALKDKKKELEGAGKIATEAIEFNFTVVSLSLTOEQKFE 360
Db 854 TLLLLAIVPIIATAGVVEKMLSGQALKDKKKELEGAGKIATEAIEFNFTVVSLSLTOEQKFE 913
QY 361 HMYAQSLQ 368
Db 915 NMIAQSLQ 922

Db 914 TWYAQSLQ 921

RESULT 5
JH0502

P-glycoprotein - rat
N:Alternate names: multidrug resistance protein mdr1b
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 02-Feb-2001
C:Accession: JH0502; S22353; S22352
R:Silverman, J.A.; Raunio, H.; Gant, T.W.; Thorgeirsson, S.S.
Gene 106, 229-236, 1991
A:Title: Cloning and characterization of a member of the rat multidrug resistance (mdr)
A:Reference number: JH0502; MUID:92039081
A:Accession: JH0502
A:Molecule type: mRNA
A:Residues: 1-1277 <STL>
A:Cross-references: GB:M62425
R:Deuchars, K.L.; Duthie, M.; Ling, V.
Biochim. Biophys. Acta 1130, 157-165, 1992
A:Title: Identification of distinct P-glycoprotein gene sequences in rat.
A:Reference number: S22351; MUID:92223089
A:Accession: S22353
A:Molecule type: DNA
A:Residues: 1212-1226, 'I', 1228-1277 <DEU>
A:Cross-references: EMBL:X61104; NID:956890; PIDN:CAA43416.1; PID:g1334219
A:Accession: S22352
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1212-1226, 'I', 1228-1270, 'SV', 1271-1277 <DE2>
A:Cross-references: EMBL:X61103; NID:956888; PIDN:CAA43415.1; PID:g1334218
C:Comment: This protein has the nucleotide binding motifs and ATP binding active transpo
C:Genetics:
A:Gene: mdr1b
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; membrane protein; nucleotide binding; P-loop
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-434/Region: nucleotide-binding motif A (P-loop)
F:550-554/Region: nucleotide-binding motif B
F:1053-1248/Domain: ATP-binding cassette homology <ABC2>
F:1070-1078/Region: nucleotide-binding motif A (P-loop)
F:1195-1199/Region: nucleotide-binding motif B
F:132/Binding site: ATP (Lys) #status predicted
F:1076/Binding site: ATP (Lys) #status predicted

Query Match 77.3%; Score 1418.5; DB 2; Length 1277;
Best Local Similarity 74.3%; Pred. No. 5.8e-106;
Matches 275; Conservative 56; Mismatches 36; Indels 3; Gaps 2;

QY 1 TSALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDGQVIVEKGNHDELMK 60
Db 557 TSALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDGQVIVEKGNHDELMK 616
QY 61 EKGIFKLVMTQTRGNEIDLENATGSKESDALEMSPKDSGLIKRRSTRSIHAPOG 120
Db 617 EKGIFKLVMTQTRGNEIEPGNNAYESQDTGASELSSEKSPLI-RRSIRSRHROD 675
QY 121 QDRKLGTKEDLNENPVVFWRIKLNSTEWPYFVVGIFCAILINGLOQPAFISRIIG 180
Db 676 QERRLSKEDVDVDPWSEFQIKLNISEWPVYVGVLCVAVNGCIQPVFAIVFSKIVG 735
QY 181 IFRDEDPETKRQISNMFSVLFLVGLIISFTFFLQGGFTFGKAGEILLTKRLRYMYFRSML 240
Db 736 VFSRDDHETKQRNCNLFSLFLVMGMISFVYFFQGGFTFGKAGEILLTKRLRYMYFRSML 795
QY 241 RQVSNFDDLLKNTTGALTTRFLANDAAQVGAIGSRSLAVITQIANLGTGIIS--LIYGW 298
Db 796 RQDISFDDHKNTGSLTTRFLASDASNVKGMSRLAVYQNVANLGTGIISLVLVYGW 855
QY 299 QLTLALLAIVPIIAIGVEMKMLSGOALKDKKEBAGKIAATEAENRPTVVSUTQEQK 358
Db 856 QLTLALLVVIPLVIGIEMKLSGOALKDKKEBAGKIAATEAENRPTVVSUTREQK 915

QY 359 FEHMYAQSLQ 368
Db 916 FETMYAQSLQ 925

RESULT 6

DVMSI
multidrug resistance protein 1 - mouse
N:Alternate names: P-glycoprotein 1
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: A33719; A25057; I57510
R:Raymond, M.; Gros, P.
Proc. Natl. Acad. Sci. U.S.A. 86, 6488-6492, 1989
A:Title: Mammalian multidrug-resistance gene: correlation of exon organization with s
A:Reference number: A33719; MUID:89367274
A:Accession: A33719
A:Molecule type: DNA
A:Residues: 1-1276 <RAY>
R:Gros, P.; Croop, J.; Housman, D.
Cell 47, 371-380, 1986
A:Title: Mammalian multidrug resistance gene: complete cDNA sequence indicates strong
A:Reference number: A25057; MUID:87028229
A:Accession: A25057
A:Molecule type: mRNA
A:Residues: 1-1276 <GRO>
A:Cross-references: GB:M14757; NID:g199100; PIDN:AAA9005.1; PID:g387426
R:Raymond, M.; Gros, P.
Mol. Cell. Biol. 10, 6036-6040, 1990
A:Title: Cell-specific activity of cis-acting regulatory elements in the promoter of
A:Reference number: I57510; MUID:91042535
A:Accession: I57510
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-21 <RES>
A:Cross-references: GB:M60348; NID:g199102; PIDN:AAA39513.1; PID:g554199
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant c
tructurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: mdr1 (pgpl)
A:Introns: 21/2; 38/3; 92/1; 111/2; 175/2; 233/3; 274/2; 332/3; 370/3; 407/3; 449/3;
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane
F:1-637,653-1276/Region: duplication
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:550-554/Region: nucleotide-binding motif B
F:1051-1247/Domain: ATP-binding cassette homology <ABC2>
F:1068-1075/Region: nucleotide-binding motif A (P-loop)
F:1194-1198/Region: nucleotide-binding motif B
F:73,91,96,103/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:432/Binding site: ATP (Lys) #status predicted
F:1074/Binding site: ATP (Lys) #status predicted

Query Match 76.3%; Score 1401.5; DB 1; Length 1276;
Best Local Similarity 72.6%; Pred. No. 1.4e-104;
Matches 267; Conservative 59; Mismatches 41; Indels 1; Gaps 1;

QY 1 TSALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDGQVIVEKGNHDELMK 60
Db 557 TSALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDGQVIVEKGNHDELMR 616
QY 61 EKGIFKLVMTQTRGNEIDLENATGSKESDALEMSPKDSGLIKRRSTRSIHAPOG 120
Db 617 EKGIFKLVMTQTRGNEIEPGNNAYQSQDTGASELSSEKSPLI-RRSIRSRHRRKD 675
QY 121 QDRKLGTKEDLNENPVVFWRIKLNSTEWPYFVVGIFCAILINGLOQPAFISRIIG 180
Db 676 QERRLSKAEVDVDPVLSFWRIKLNSTEWPVYVGVLCVAVNGCIQPVFAIVFSRIG 735
QY 181 IFRDEDPETKRQISNMFSVLFLVGLIISFTFFLQGGFTFGKAGEILLTKRLRYMYFRSML 240
Db 736 VFSRDDHETKQRNCNLFSLFLVMGLISFVYFFQGGFTFGKAGEILLTKRLRYMYFRSML 795

F:432/Binding site: ATP (lys) #status predicted
F:1072/Binding site: ATP (Lys) #status predicted

Query Match 66.5%; Score 1220.5; DB 1; Length 1276;
Best Local Similarity 67.6%; Pred. No. 5.3e-90;
Matches 248; Conservative 49; Mismatches 67; Indels 3; Gaps 3;

Qy 1 TSALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGQVIVEQSHSELK 60
Db 557 TSALDTESEAVQALDKAREGRTTIVIAHRLSTVRNADVIAGFDGQVIVEQSHSELK 616
Qy 61 EKGIFYKLVMTQTRGNEIDLENATGESKSDALEMSPKDSGLIKRRSTRRSHAPQG 120
Db 617 KEGYFRLVNMQTSGQILSEFEVELSDEKAAGVAP-NGWKARIFRNSTKSLKSSRA 674
Qy 121 QDRKLGTK-EDLNENPVPSFWRIKLKLNSTENWYFVVGIFCAIINGLOPAFSIILSEMI 180
Db 675 QNRLDEETNELDANPPVPSFLKVLKLNKTEWYFVVGITCAIANGALQPAFSIILSEMI 734
Qy 181 IFTREDDEPTKQISNMFSVLFLVGLIISFTFFLQGGTFGKAGEILTKRLRYMVFRSML 240
Db 735 IFPGDD-AVKQKCNMFSVLFLGLVLSFTFFLQGGTFGKAGEILTKRLRSMFAKML 793
Qy 241 RDQVSWFDDKNTGALTTRLANDAAQVKGATGSLRAVITONIANLGTGIIISLIYGMQ 300
Db 794 RDQMSWFDHKNSTGALSTRLATDAQVQCATGTRLALIAQNTANLGTGIIISFIYGMQ 853
Qy 301 TLLLLAIVPIIAIAGVVENMKMLSGQALKDKKELEGAGKIATEAIENFRTVVSLSLTOEOKF 360
Db 854 TLLLSVVPFIAVAGIVEMKMLAGNAKRDKKEAAGKIATEAIENIRTVVSLTOERKF 913
Qy 361 HMYAQSL 367
Db 914 SMYVEKL 920

RESULT 9
S41646
P-glycoprotein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 02-Feb-2001
C:Accession: S41646; S22354
R:Brown, P.C.; Thorgerlsson, S.S.; Silverman, J.A.
Nucleic Acids Res. 21, 3885-3891, 1993
A:Title: Cloning and regulation of the rat mdr2 gene.
A:Reference number: S41646; MUID:93376516
A:Accession: S41646
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1278 <BRO>
A:Cross-references: EMBL:L15079; NID:g310192; PIDN:AAA02937.1; PID:g310193
R:Deuchars, K.L.; Duthie, M.; Ling, V.
Biochim. Biophys. Acta 1130, 157-165, 1992
A:Title: Identification of distinct p-glycoprotein gene sequences in rat.
A:Reference number: S22351; MUID:922223089
A:Accession: S22354
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1211-1278 <DEU>
A:Cross-references: EMBL:X61105; NID:g56892; PIDN:CAA43417.1; PID:g1334220
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
F:409-603/Domain: ATP-binding cassette homology <ABC1>
F:426-433/Region: nucleotide-binding motif A (P-loop)
F:1051-1247/Domain: ATP-binding cassette homology <ABC2>
F:1068-1075/Region: nucleotide-binding motif A (P-loop)

Query Match 65.8%; Score 1207.5; DB 2; Length 1278;
Best Local Similarity 66.8%; Pred. No. 5.9e-89;
Matches 246; Conservative 49; Mismatches 70; Indels 3; Gaps 3;

Qy 1 TSALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGQVIVEQSHSELK 60
Db 557 TSALDTESEAVQALDKAREGRTTIVIAHRLSTVRNADVIAGFDGQVIVEQSHSELK 616

6

Db 557 TSALDTESEAVQALDKAREGRTTIVIAHRLSTVRNADVIAGFDGQVIVEQSHSELK 616
Qy 61 EKGIFYKLVMTQTRGNEIDLENATGESKSDALEMSPKDSGLIKRRSTRRSHAPQG 120
Db 617 KEGYFRLVNMQTSGQILSEFEVELSDEKAAGVAP-NGWKARIFRNSTKSLKSSRA 675
Qy 121 QDRKLGTK-EDLNENPVPSFWRIKLKLNSTENWYFVVGIFCAIINGLOPAFSIILSEMI 179
Db 676 QNRLDEETNELDANPPVPSFLKVLKLNKTEWYFVVGITCAIANGALQPAFSIILSEMI 735
Qy 180 GIFTREDDEPTKQISNMFSVLFLVGLIISFTFFLQGGTFGKAGEILTKRLRYMVFRSM 239
Db 736 AIFPGDD-TVQKQCNMFSVLFLGLVLSFTFFLQGGTFGKAGEILTKRLRSMFAKML 794
Qy 240 RDQVSWFDDKNTGALTTRLANDAAQVKGATGSLRAVITONIANLGTGIIISLIYGMQ 299
Db 795 RDQMSWFDHKNSTGALSTRLATDAQVQCATGTRLALIAQNTANLGTGIIISFIYGMQ 854
Qy 300 TLLLLAIVPIIAIAGVVENMKMLSGQALKDKKELEGAGKIATEAIENFRTVVSLSLTOEOKF 359
Db 855 TLLLSVVPFIAVAGIVEMKMLAGNAKRDKKEAAGKIATEAIENIRTVVSLTOERKF 914
Qy 360 HMYAQSL 367
Db 915 SMYVEKL 922

Query Match 65.6%; Score 1204.5; DB 2; Length 1281;
Best Local Similarity 66.6%; Pred. No. 1e-88;
Matches 245; Conservative 48; Mismatches 72; Indels 3; Gaps 3;

Qy 1 TSALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGQVIVEQSHSELK 60
Db 560 TSALDTESEAVQALDKAREGRTTIVIAHRLSTVRNADVIAGFDGQVIVEQSHSELK 619
Qy 61 EKGIFYKLVMTQTRGNEIDLENATGESKSDALEMSPKDSGLIKRRSTRRSHAPQG 120
Db 620 KEGYFRLVNMQTSGQILSEFEVELSDEKAAGVAP-NGWKARIFRNSTKSLKSSRA 678
Qy 121 QDRKLGTK-EDLNENPVPSFWRIKLKLNSTENWYFVVGIFCAIINGLOPAFSIILSEMI 179
Db 679 HHRLDVDADELNANPPVPSFLKVLKLNKTEWYFVVGITCAIANGALQPAFSIILSEMI 738
Qy 180 GIFTREDDEPTKQISNMFSVLFLVGLIISFTFFLQGGTFGKAGEILTKRLRYMVFRSM 239
Db 739 AIFPGDD-AVKQKCNMFSVLFLGLVLSFTFFLQGGTFGKAGEILTKRLRSMFAKML 797
Qy 240 RDQVSWFDDKNTGALTTRLANDAAQVKGATGSLRAVITONIANLGTGIIISLIYGMQ 299
Db 798 RDQMSWFDHKNSTGALSTRLATDAQVQCATGTRLALIAQNTANLGTGIIISFIYGMQ 857

```

Qy 300 L T L L L A I V P I I A I G V V E K M L S G A L K D K K E L S G A K I A T E A I E N P R T V V S L T Q R K F 359
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 858 L T L L L S V P F I A V S G I V E K M L A G N A K R D K K A L E A A G K I A T E A I E N I R T V V S L T Q R K F 917
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 360 E H M Y A Q S L 367
      |||::|
Db 918 E S M T V E K L 925

RESULT 11
S55692
multidrug resistance protein homolog (mdr) - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 02-Feb-2001
C:Accession: S55692
R:Castillo, G.; Shen, H.J.; Horwitz, S.B.
Biochim. Biophys. Acta 1262, 113-123, 1995
A:Title: A homologue of the mammalian multidrug resistance gene (mdr) is fu
A:Reference number: S53692; MUID:95322451
A:Accession: S55692
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1287 <AS>
A:Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; nucleotide binding; P-loop
F:420-614/Domain: ATP-binding cassette homology <ABC1>
F:437-444/Region: nucleotide-binding motif A (P-loop)
F:1062-1258/Domain: ATP-binding cassette homology <ABC2>
F:1079-1086/Region: nucleotide-binding motif A (P-loop)

Query Match 62.5%; Score 1148; DB 2; Length 1287;
Best Local Similarity 62.5%; Pred. No. 3.7e-84;
Matches 235; Conservative 61; Mismatches 62; Indels 18; Gaps

Qy 1 T S A L D T S E A V V Q A L D K A R K G R T T I V I A H R L S T V R N A D V I A G F D D G V I V E K G N H D E L M K 60
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 568 T S A L D T S E A V V Q A L D K A R E G R T T I V A H R L S T I R N A N A I A G F D G V I V E Q G S H K E L M E 627
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 61 E K G I Y F K L V M Q T - - - - R G N E I D L N A T G E S K E S D A L E M S P K D S S S I I K R R S P R R S I H 116
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 628 R G G V I F N L T L Q T V E T S K D T E D L E T H I Y E K - - - - - I P V T H S N L V R K R S S R N T I K 680
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 117 A - - P Q G D R K L G T K E - D L N E N P V P S F W R I L K I N S T E W P Y F V G I F C A I I N G L Q P A F S I 173
      :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 681 S K V P E T D E K V D E E K K E S G P P V S F F K V M L K N K E W P Y F V G V I C A M I N G A T Q P A F A I 740
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 174 I F S R I I G I F T R D E D P E T K - R Q I S N M S P L F L V L G I I S F T F F L Q G F T F G K A G E I L T K R L R 232
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 741 I F S R I I G V F A - - - G P V S Q M R S E S S M Y S L L F A I L G V S F T F F L Q R F T F G K A G E I L T M R L R 797
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 233 Y V F S R M L R Q D V S F D D L K N T T G A L T T R L A N D A A O V K G A I S R L A V I Q N T A N I L G T I I I 292
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 798 L G S F K S M L R Q B I G F W D D S K N T G A L T T R L A T D A S Q V G A T R L L A A Q N V A N L G T A I I I 857
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 293 S L I Y G W L T L L L A I V P I I A I G V V E K M L S G A L K D K K E L S G A K I A T E A I E N P R T V V S 352
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 858 S F I Y G W L T L L L A I V P I I A V I A A G L V E K M F A G H A K D K K E L S G A K I S T D A V L I R T V V S 917
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy 353 L T Q E Q K F E H M Y A Q S L 368
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 918 L T R E K K F E A M Y E K S L E 933

RESULT 12
DWHY2C
multidrug resistance protein 2 - Chinese hamster (fragment)
N:Alternate names: P-glycoprotein pgp2
C:Species: Cricetus griseus (Chinese hamster)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 19-Jan-2001
C:Accession: B37126
R:Endicott, J.A.; Juranka, P.F.; Sarangi, F.; Gerlach, J.H.; Deuchars, K.L.
Mol. Cell. Biol. 7, 4075-4081, 1987

```

[illegible]

A>Title: Simultaneous expression of two p-glycoprotein genes in drug-sensitive Chinese
A:Reference number: A27126; MUID:88122132
A:Accession: B27126
A:Molecule type: mRNA
A:Residues: 1-655 <END>
A:Cross-references: GB:ML7896; NID:g191161; PIDN:AAA37007.1; PID:g387054
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant C.
structurally and functionally unrelated lipophilic antitumor drugs.
C:Genetics:
A:Gene: pgp2
C:Superfamily: multidrug resistance protein; ATP-binding cassette homolog
C:Keywords: App; duplication; glycoprotein; nucleotide binding; P-loop; transmembrane
F:430-626/Domain: ATP-binding cassette homology <ABC2>
F:447-454/Region: nucleotide-binding motif A (P-loop)
F:573-577/Region: nucleotide-binding motif B
F:453/Binding site: ATP (Lys) #status predicted

Query Match 58.4%; Score 1071.5; DB 1; Length 655;
Best Local Similarity 68.9%; Pred. No. 2.3e-78;
Matches 208; Conservative 50; Mismatches 43; Indels 1; Gaps 1;

Qy 67 KLVTMTGRGNDILENATGESKSDALEMSPKDSGLIKRRSTRSIHAPQGDRKLQ 126
||| ||||| :
Db 2 RLVMQTGRNEVELGSEADGSQSDTIAELSTSEEFSPSV-RKSTCRSICGSDQERRVS 60
||| ||||| :
Qy 127 TKEDLENVPVPSFWIRILKLNSTEPPYVVGIFCAINGLQPAFSIFSRIGITRDE 186
||| : : ||| ||||| :
Db 61 VKEAQDEDPLVPSFWGI LKLNITEWPYLGVGLCAVINGCMQPVFSIVFGIIGVFTRDD 120
||| : : ||| ||||| :
Qy 187 DPEYKRQISNFWSVLFLVGLGISITFFLOGFTFGKAGEILT KRLRYMVFSMLRQDVSW 246
||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 121 DPKTQQKNCLFSLFFLVGMGICFYTFYFOGFTFGKAGEILT KRLRYMVFSMLRQDLISW 180
||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Qy 247 FDDLKNTTGALTTLANDAAOVKGASRLAVITONTIANLTGTIIISLIYGWQLTLILA 306
||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 181 FDDRHNSTGALTTLASDANVKAMSSRLAGITQNVA NLGTGIIISLYGWQLTLILVV 240
||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Qy 307 IVPIITAAGVVMKMSGLAKDKKELEGAKATEAIENFRTVVSLTQEOKFHEHTAQS 366
||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 241 IAPLLISGMEMKVLSGALKDKKELEVSCKIATEIENFRTVVSLTREQKFENMYAQS 300
||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Qy 367 IQ 368
||
Db 301 LQ 302

RESULT 13
T42228
P-glycoprotein sister - rat
N:Alternate names: multidrug resistance protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Feb-2001
C:Accession: T42228
R:Childs, S.J.; Yeh, R.L.; Hui, D.; Ling, W.
submitted to the EMBL Data Library, June 1997
A:Description: Taxol resistance mediated by the liver-specific Sister gene of P-glyco
protein
A:Reference number: Z22088
A:Accession: T42228
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1321 <CHI>
A:Cross-references: EMBL:AF010597; NID:g3273483; PID:g3273484; PIDN:AAC24753.1
C:Genetics:
A:Gene: spgp
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
C:Keywords: glycoprotein; membrane protein; nucleotide binding; P-loop
F:455-462/Region: nucleotide-binding motif A (P-loop)

Query Match 42.6%; Score 782.5; DB 2; Length 1321;
Best Local Similarity 43.5%; Pred. No. 1.1e-54;
Matches 167; Conservative 76; Mismatches 124; Indels 17; Gaps 6;

Qy 1 TSALDTESAVVOVALDKARKGRITVTAHLRUSTVRNADVTAGDDGVIVEKGNDELMMK 60

Db 586 TSALDNESARVQALNKIQHGHHTIIISVAHRLSTVRAADVILGFHGVAVERTHEELE 645
QY 61 EKGIFKLVMTQTRNGEIDLENA-TGESKSEDALEM-----SPKDSGSLIKRRSTR- 113
Db 646 RKGIVFMLVTLQSGDNNAHRETSIMGKDATEGGTLETFSRGYSRDLRASIRQRKSQ 705
QY 114 --SIHAP--QGDRKLGTKED-----LNENVPVPSFWRILKLNSTWPFVVGIFCAIN 164
Db 706 SLLTHDPLAVADHKSSYKSDNDVLEVEEPAPVRRLKYNIPWHYILVGLSAAIN 765
QY 165 GGLQPAFSIIFSRIGITFRDEDPETKQISNMFSVLFLVLGIISFTIFLQGTFFGKAG 224
Db 766 GAVTPIYSLFLSOLLGTFSLLDKEQORSEIHSW-CLFFVILGCVSIFTOFLQGYTFKSG 824
QY 225 EILTRLRYMVFRLSRQDYSWFDLKNNTGALTTRLANDAAQVKAIGSLRAVITQNTA 284
Db 825 ELLTKRLRKFEGKAMLGQDIFGWFDLLRNNPGVLTTRLATDASQVQATGSQVGMVNST 884
QY 285 NLGTGIIISLIYGWOLTLALLAIVPIAIAAGVVMKMLSGQALKDKEGAGKIATEAI 344
Db 885 NIITAAALIAFFSWKLSLIITIFFPFLALSGAVQTKMLTGFASQDKQALEKAGQITSEAL 944
QY 345 ENFRTVVSLTQEQKFEHMYAQSILQ 368
Db 945 SNIRTVAGIGVEGRFIKAFVELQ 968

RESULT 14
T42842
bile salt transport protein, ATP-dependent - rat
N:Alternate names: bile salt export pump, sister of P-glycoprotein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-May-2000
C:Accession: T42842
R:Gerloff, T.; Stieger, B.; Hagenbuch, B.; Madon, J.; Landmann, L.; Roth, J.; Hofmann, A.
J. Biol. Chem. 273, 10046-10050, 1998
A:Title: The sister of P-glycoprotein represents the canalicular bile salt export pump
A:Reference number: P22272; MUID:98212048
A:Accession: T42842
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1321 <GER>
A:Cross-references: EMBL:U69487; NID:g3075421; PID:g3075422; PIDN:AAC40084.1
A:Experimental source: strain Sprague-Dawley; liver
C:Genetics:
A:Gene: spgp
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; membrane protein; P-loop

Query Match 42.6%; Score 782.5; DB 2; Length 1321;
Best Local Similarity 43.5%; Pred. No. 1.le-54;
Matches 167; Conservative 76; Mismatches 124; Indels 17; Gaps 6;

QY 1 TSALDTESEAVVQVALDKARKGRITIVIAHRLSTVRNADVIAGFDDGVIYVEKGNHDELMK 60
Db 586 TSALDNESARVQALNKIQHGHHTIIISVAHRLSTVRAADVILGFHGVAVERTHEELE 645
QY 61 EKGIFKLVMTQTRNGEIDLENA-TGESKSEDALEM-----SPKDSGSLIKRRSTR- 113
Db 646 RKGIVFMLVTLQSGDNNAHRETSIMGKDATEGGTLETFSRGYSRDLRASIRQRKSQ 705
QY 114 --SIHAP--QGDRKLGTKED-----LNENVPVPSFWRILKLNSTWPFVVGIFCAIN 164
Db 706 SLLTHDPLAVADHKSSYKSDNDVLEVEEPAPVRRLKYNIPWHYILVGLSAAIN 765
QY 165 GGLQPAFSIIFSRIGITFRDEDPETKQISNMFSVLFLVLGIISFTIFLQGTFFGKAG 224
Db 766 GAVTPIYSLFLSOLLGTFSLLDKEQORSEIHSW-CLFFVILGCVSIFTOFLQGYTFKSG 824
QY 225 EILTRLRYMVFRLSRQDYSWFDLKNNTGALTTRLANDAAQVKAIGSLRAVITQNTA 284
Db 825 ELLTKRLRKFEGKAMLGQDIFGWFDLLRNNPGVLTTRLATDASQVQATGSQVGMVNST 884

QY 285 NLGTGIIISLIYGWOLTLALLAIVPIAIAAGVVMKMLSGQALKDKEGAGKIATEAI 344
Db 885 NIITAAALIAFFSWKLSLIITIFFPFLALSGAVQTKMLTGFASQDKQALEKAGQITSEAL 944
QY 345 ENFRTVVSLTQEQKFEHMYAQSILQ 368
Db 945 SNIRTVAGIGVEGRFIKAFVELQ 968

RESULT 15
T31073
multidrug resistance p-glycoprotein - nematode (Haemonchus contortus)
C:Species: Haemonchus contortus
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T31073
R:Xu, M.; Molento, M.; Blackhall, W.; Ribeiro, P.; Beech, R.; Prichard, R.
Mol. Biochem. Parasitol. 91, 327-335, 1998
A:Title: Ivermectin resistance in nematodes may be caused by alteration of P-glycopro
A:Reference number: Z20968; MUID:98226176
A:Accession: T31073
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1275 <XOM>
A:Cross-references: EMBL:AF003908; NID:g3057041; PID:g3057042; PIDN:AAC38987.1
A:Experimental source: developmental stage adult
C:Genetics:
A:Gene: PGP-A
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology

Query Match 36.9%; Score 677.5; DB 2; Length 1275;
Best Local Similarity 37.1%; Pred. No. 3e-46;
Matches 137; Conservative 98; Mismatches 127; Indels 7; Gaps 6;
QY 1 TSALDTESEAVVQVALDKARKGRITIVIAHRLSTVRNADVIAGFDDGVIYVEKGNHDELMK 60
Db 559 TSALDTEASEIVQALEKAQKGRITIVIAHRLSTIRNVQDIFVFKNGTIVEQGTTHAELMN 618
QY 61 EKGIFKLVMTQTRNGEIDLENAIGESKSEDALEMSPKDSGSLIKRRSTRRSIHA-PQ 119
Db 619 KRGVFFEMTQAVLRQKEKEEVLDSDAESDVVSPDIALPHLSLSRKESTRSAISAVPS 678
QY 120 QGDRKLGTKEDLNENVPVPSFWRILKLNSTWPFVVGIFCAIINGLQPAFSIIFSRII 179
Db 679 VRSMQI-EMEDLRAPKTPMS--KIFYFNDRKMGYFILGLACIITGTVTFAVLYAQII 735
QY 180 GIFTRDEDPETKQISNMFSVLFLVLGIISFTIFLQGTFFGKAGEILTTLRLRYMVFPSM 239
Db 736 QVYSEPDV-QMKGHVL-FWCGAFIVILGVHAFAPFFSAICLGRGCEALTTLKLRFEAFKNL 793
QY 240 LRQDVSWFDDLLKNNTGALTTRLANDAAQVKAIGSLRAVITONIANLGTIIISLIYGWQ 299
Db 794 LRQNVGFDDIRHGTGKLCITRFATDAPNVR-YVETRLPGVLSVVTIGALVIGFIFGWQ 852
QY 300 LTLLALLAIVPIAIAAGVVMKMLSGQALKDKEGAGKIATEAIENFRVVSILTQEQKF 359
Db 853 LALLTMVMPVLIIGSGYFEMRMQFGKKMRDTELEEEAGKVASQAVENIRTVHAINROEQF 912
QY 360 EHMVAQSILQ 368
Db 913 HFMYCEYLK 921

Search completed: November 6, 2002, 18:44:42
Job time : 10.7636 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 18:30:44 ; Search time 3.01584 Seconds
(without alignments)
4724.646 Million cell updates/sec

Title: US-09-672-725C-6

Perfect score: 1836

Sequence: 1 TSALDTSEAVVQVALDKR.....TVVSLTQEQKFEHMYAQSLQ 368

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1672	91.1	1280	1 MDR1_HUMAN	P08183 homo sapien
2	1589	86.5	1276	1 MDR1_CRIGR	P21448 cricetus
3	1584	86.3	1276	1 MDR3_MOUSE	P21447 mus musculus
4	1418.5	77.3	1277	1 MDR1_RAT	P43245 rattus norv
5	1401.5	76.3	1276	1 MDR1_MOUSE	P06795 mus musculus
6	1366.5	74.4	1276	1 MDR2_CRIGR	P21449 cricetus
7	1248.5	68.0	1279	1 MDR3_HUMAN	P21439 homo sapien
8	1220.5	66.5	1276	1 MDR2_MOUSE	P21440 mus musculus
9	1207.5	65.8	1278	1 MDR2_RAT	Q08201 rattus norv
10	1204.5	65.6	1281	1 MDR3_CRIGR	P23174 cricetus
11	803.5	43.8	1321	1 AB11_RABIT	Q9n0v3 oryctolagus
12	794.5	43.3	1321	1 AB11_MOUSE	Q9gy30 mus musculus
13	788.5	42.9	1321	1 AB11_HUMAN	Q95342 homo sapien
14	782.5	42.6	1321	1 AB11_RAT	O70127 rattus norv
15	605.5	33.0	1302	1 MDR4_DROME	Q00449 drosophila
16	597.5	32.5	1302	1 MDR5_DROME	Q00748 drosophila
17	589.5	32.1	1321	1 MDR1_CABEL	P34712 caenorhabdi
18	490	26.7	1254	1 MDR3_CABEL	P34713 caenorhabdi
19	490	26.7	1362	1 PMD1_SCHPO	P36619 schizosacch
20	426.5	23.2	1280	1 MDR1_LETEN	Q06034 leishmania
21	301.5	16.4	1419	1 MDR_PLAFF	P13568 plasmodium
22	261	14.2	1290	1 STE6_YEAST	P12866 saccharomyc
23	258	14.1	1336	1 MAM1_SCHPO	P78966 schizosacch
24	253	13.8	1323	1 HST6_CANAL	P33706 candida alb
25	215	11.7	582	1 MSBA_ECOLI	P27299 escherichia
26	200	10.9	587	1 MSBA_HAEIN	P44407 haemophilus
27	199.5	10.9	735	1 ABC8_HUMAN	Q9nut2 homo sapien
28	195	10.6	1436	1 MRP5_RAT	O9qym0 rattus norv
29	194	10.6	598	1 Y288_THEMEA	Q9wyc4 thermotoga
30	193.5	10.5	1437	1 MRP5_HUMAN	O15440 homo sapien
31	193	10.5	1436	1 MRP5_MOUSE	Q9rlx5 mus musculus
32	192.5	10.5	1545	1 MRP2_HUMAN	Q92887 homo sapien
33	190.5	10.4	1564	1 MRP2_RABIT	Q28689 oryctolagus

RESULT 1

ID	MDR1_HUMAN	STANDARD;	PRT; 1280 AA.
AC	P08183; Q12755; Q14812;		
DT	01-AUG-1988 (Rel. 08, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	18-OCT-2001 (Rel. 40, Last annotation update)		
DE	Multidrug resistance protein 1 (P-glycoprotein 1).		
GN	ABCBI OR PGY1 OR MDR1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=87028230; PubMed=2876781;		
RA	Chen C.-J., Chin J.E., Ueda K., Clark D.P., Pastan I., Gottesman M.M.,		
RA	Roninson I.B.;		
RT	"Internal duplication and homology with bacterial transport proteins		
RT	in the mdr1 (P-glycoprotein) gene from multidrug-resistant human		
RT	cells.";		
RL	Cell 47:381-389(1986).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90094448; PubMed=1967175;		
RA	Chen C.-J., Clark D.P., Ueda K., Pastan I., Gottesman M.M.,		
RA	Roninson I.B.;		
RT	"Genomic organization of the human multidrug resistance (MDR1) gene		
RT	and origin of P-glycoproteins.";		
RL	J. Biol. Chem. 265:506-514(1990).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97190336; PubMed=9038218;		
RA	Chen G., Duran G.E., Steger K.A., Lacayo N.J., Jaffrezou J.P.,		
RA	Dumontet C., Sikic B.I.;		
RT	"Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein,		
RT	altered phenotype, and resistance to cyclosporins.";		
RL	J. Biol. Chem. 272:5974-5982(1997).		
RN	[4]		
RP	SEQUENCE OF 1-234 FROM N.A.		
RA	Smith A., Beck C., Gibson A.;		
RA	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.		
RN	[5]		
RP	SEQUENCE OF 178-215 AND 800-856 FROM N.A.		
RX	MEDLINE=90290529; PubMed=1972623;		
RA	Gekeler V., Weger S., Probst H.;		
RT	"mdr1/P-glycoprotein gene segments analyzed from various human		
RT	leukemic cell lines exhibiting different multidrug resistance		
RT	profiles.";		
RL	Biochem. Biophys. Res. Commun. 169:796-802(1990).		
RN	[6]		
RP	SEQUENCE OF 1-23 FROM N.A.		
RA	Kioka N., Tsubota J., Kakehi Y., Komano T., Gottesman M.M.,		
RA	Pastan I., Uedak K.;		
RL	Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.		
CC	!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED		

P97046 lactococcus
P18768 agrobacteri
Q02592 schizosacch
Q9chl8 lactococcus
P35598 streptococc
Q9np58 homo sapien
Q11047 mycobacteri
P54719 bacillus su
P23702 actinobacil
Q04473 actinobacil
Q03519 homo sapien
P97398 candida alb

ALIGNMENTS

34 185 10.1 584 1 LMRA_LACLC
35 185 10.1 588 1 CHVA_AGRU
36 184.5 10.0 830 1 HMT1_SCHPO
37 184 10.0 584 1 LMRA_LACLA
38 181 9.9 583 1 EX98_STRPN
39 179.5 9.8 842 1 ABC6_HUMAN
40 176 9.6 631 1 YC72_MYCTU
41 175 9.5 604 1 YFIC_BACSU
42 174.5 9.5 707 1 HLYB_ACTAC
43 173.5 9.4 711 1 RT3B_ACTPL
44 172.5 9.4 686 1 TAP2_HUMAN
45 172 9.4 685 1 MDLI_CANAL

CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, SMALL INTESTINE
 CC AND BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M14758; AAA59575.1; -
 CC EMBL; M29447; AAA59576.1; -
 CC EMBL; M29424; AAA59576.1; JOINED.
 CC EMBL; M29425; AAA59576.1; JOINED.
 CC EMBL; M29426; AAA59576.1; JOINED.
 CC EMBL; M29427; AAA59576.1; JOINED.
 CC EMBL; M29428; AAA59576.1; JOINED.
 CC EMBL; M29429; AAA59576.1; JOINED.
 CC EMBL; M29430; AAA59576.1; JOINED.
 CC EMBL; M29431; AAA59576.1; JOINED.
 CC EMBL; M29432; AAA59576.1; JOINED.
 CC EMBL; M29433; AAA59576.1; JOINED.
 CC EMBL; M29434; AAA59576.1; JOINED.
 CC EMBL; M29435; AAA59576.1; JOINED.
 CC EMBL; M29436; AAA59576.1; JOINED.
 CC EMBL; M29437; AAA59576.1; JOINED.
 CC EMBL; M29438; AAA59576.1; JOINED.
 CC EMBL; M29439; AAA59576.1; JOINED.
 CC EMBL; M29440; AAA59576.1; JOINED.
 CC EMBL; M29441; AAA59576.1; JOINED.
 CC EMBL; M29442; AAA59576.1; JOINED.
 CC EMBL; M29443; AAA59576.1; JOINED.
 CC EMBL; M29444; AAA59576.1; JOINED.
 CC EMBL; M29445; AAA59576.1; JOINED.
 CC EMBL; M29446; AAA59576.1; JOINED.
 CC EMBL; AF016535; AAB69423.1; -
 CC EMBL; AC002457; AAC82531.1; -
 CC EMBL; M37724; AAA88047.1; -
 CC EMBL; M37725; AAA88048.1; -
 CC EMBL; X58723; CAA41558.1; -
 CC PIR; A25059; DVH01.
 CC PIR; A34914; A34914.
 CC MIM; 171050; -
 CC InterPro: IPR003593; AAA.
 CC InterPro: IPR001140; ABC_transporter_tmern.
 CC InterPro: IPR003439; ABC_transporter.
 CC InterPro: IPR001687; ATP_GTP_A.
 CC Pfam: PF00564; ABC_membrane; 2.
 CC Pfam: PF00005; ABC_tran; 2.
 CC SMART; SM00382; AAA; 2.
 CC PROSITE; PS00211; ABC_TRANSPORTER; 2.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
 KW Multigene family.
 FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 52 72 POTENTIAL.
 FT TRANSMEM 120 140 POTENTIAL.
 FT TRANSMEM 189 209 POTENTIAL.
 FT TRANSMEM 216 236 POTENTIAL.
 FT TRANSMEM 297 317 POTENTIAL.
 FT TRANSMEM 326 346 POTENTIAL.
 FT DOMAIN 347 710 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 711 731 POTENTIAL.
 FT TRANSMEM 757 777 POTENTIAL.
 FT TRANSMEM 833 853 POTENTIAL.
 FT TRANSMEM 854 874 POTENTIAL.
 FT TRANSMEM 937 957 POTENTIAL.
 FT TRANSMEM 974 994 POTENTIAL.
 FT DOMAIN 995 1280 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 427 434 ATP (BY SIMILARITY).

FT NP_BIND 1070 1077
 FT REPEAT 638 1280
 FT CARBOHYD 91
 FT CARBOHYD 94
 FT CARBOHYD 99
 FT CARBOHYD 99
 FT CONFLICT 23 23
 FT CONFLICT 185 185
 FT CONFLICT 336 336
 FT CONFLICT 412 412
 FT CONFLICT 438 438
 SQ SEQUENCE 1280 AA; 141462 MW; ABIC279531F43675 CRC64;
 Query Match 91.1%; Score 1672; DB 1; Length 1280;
 Best Local Similarity 90.5%; Pred. No. 5.6e-117;
 Matches 333; Conservative 16; Mismatches 19; Indels 0; Gaps 0;
 QY 1 TSALDTESEAVVQVALDKARKGRITIVIAHRLSTVRNADVIAGFDGVIIVEKGNHDELMK 60
 DB 558 TSALDTESEAVVQVALDKARKGRITIVIAHRLSTVRNADVIAGFDGVIIVEKGNHDELMK 617
 QY 61 EKIYFKLVTMOTRGNEIDLENATGESKSDALEMSPKDSGLIKRRSTRSIIHAPOG 120
 DB 618 EKIYFKLVTMOTRGNEIDLENATGESKSDALEMSPKDSGLIKRRSTRSIIHAPOG 677
 QY 121 QDRKLGTKEDLNENPVPSFWIRILKLNSTWPFVVGIFCAIINGGLQPAFISFIISRIIG 180
 DB 678 QDRKLGTKEDLNENPVPSFWIRILKLNSTWPFVVGIFCAIINGGLQPAFISFIISRIIG 737
 QY 181 IFTREDPEYKROISNMFSVLFLVGLIISITITFFLOGTFEGKAGEILTKRLRYWVFRSML 240
 DB 738 VFTRIDDPETKRONSLFLFLALGIISITITFFLOGTFEGKAGEILTKRLRYWVFRSML 797
 QY 241 RDVSWFDDKNTTGALTTRLANDAAOVKAIGSRSLAVITONTANLGTGIIISLIYQWOL 300
 DB 798 RDVSWFDDKNTTGALTTRLANDAAOVKAIGSRSLAVITONTANLGTGIIISLIYQWOL 857
 QY 301 TLLLLAIPIIAIAGVVMKMLSGQALKDKKELEGAGKIATEAENFRVTVSLTQEQKFE 360
 DB 858 TLLLLAIPIIAIAGVVMKMLSGQALKDKKELEGAGKIATEAENFRVTVSLTQEQKFE 917
 QY 361 HMYAQSLQ 368
 DB 918 HMYAQSLQ 925
 RESULT 2
 MDRL_CRIGR STANDARD; PRT; 1276 AA.
 ID MDRL_CRIGR
 AC P21448;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Multidrug resistance protein 1 (P-glycoprotein 1).
 GN ABCB1 OR PGV1 OR PGPI.
 OS Cricetus griseus (Chinese hamster).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetus.
 OC NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92135896; PubMed=1685679;
 RA Endicott J.A., Sarangi F., Ling V.;
 RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
 RT gene family";
 RL DNA Seq. 2:89-101(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91154265; PubMed=1671863;
 RA Devine S.E., Hussain A., Davide J.P., Melera P.W.;
 RT "Full length and alternatively spliced papl transcripts in multidrug-
 RT resistant Chinese hamster lung cells.";

```
RL J. Biol. Chem. 266:4545-4555(1991).
RN [3]
RP SEQUENCE OF 706-1276 FROM N.A.
RX MEDLINE=88122132; PubMed=2893325;
RA Endicott J.A., Juranka P.F., Sarangi F., Gerlach J.H., Deuchars K.L.,
RA Ling V.;
RT "Simultaneous expression of two P-glycoprotein genes in
RT drug-sensitive Chinese hamster ovary cells.";
RL Mol. Cell. Biol. 7:4075-4081(1987).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT
CC CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
CC CANNOT.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60040; AAA68883.1; -.
DR EMBL; M59253; AAA37004.1; -.
DR EMBL; M17897; AAA37006.1; -.
DR FIR; A38696; DVHV1C.
DR InterPro; IPR0013593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmam.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
FT DOMAIN 1 50 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 213 233 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 344 707 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 708 728 POTENTIAL.
FT TRANSMEM 754 774 POTENTIAL.
FT TRANSMEM 830 850 POTENTIAL.
FT TRANSMEM 851 871 POTENTIAL.
FT TRANSMEM 934 954 POTENTIAL.
FT TRANSMEM 971 991 POTENTIAL.
FT DOMAIN 992 1276 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 424 431 ATP (POTENTIAL).
FT NP_BIND 1067 1074 ATP (POTENTIAL).
FT REPEAT 1 635
FT REPEAT 636 1276
FT CONFLICT 338 339 GA -> AP (IN REF. 2).
SQ SEQUENCE 1276 AA; 140925 MW; 44F3F92A186B4DFF CRC64;
```

```
Query Match
Best Local Similarity 86.5%; Score 1589; DB 1; Length 1276;
Matches 313; Conservative 29; Mismatches 26; Indels 0; Gaps 0;

QY 1 TSALDTESEAVQVALDKARKGRITIVIAHRLSTVRNADVIAGDDGVIVKGNHDELMK 60
Db 555 TSALDTESEAVQVALDKARKGRITIVIAHRLSTVRNADVIAGDDGVIVKGNHDELMR 614
QY 61 EKGIFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRSTRRSIHAPQG 120
Db 615 EKGIFKLVMTQTAGNEIELGNEVGEKNEIDNLDMSKDSASSLIRRRSTRSIRGPHD 674
```

```
QY 121 QDRKLGTKEDLNENVPVSWFWRILKLNSTWPEVVGIFCAIINGLOPAFSIIFSRIG 180
Db 675 QDRKLSKEALDEDVPPISFWRIKLNSSEPPYVVGIFCAIYNGALOPAFSIFSKVG 734
QY 181 IFTREDEPDKROISNMFSVLFLVIGISITFTFFLQGTFTGKAGEIILTKRLRYMVFRSML 240
Db 735 VFTRNTDDETKRHSNLSFLILFLIGVISITFTFFLQGTFTGKAGEIILTKRLRYMVFRSML 794
QY 241 RDVSWPDDLKNTTTCALTTRLANDAAOVKGAIGSRSLAVITONTANLGTGIIISLIYQWL 300
Db 795 RDVSWFDPNPKNTTTCALTTRLANDAGVKGATGARLAVITONTANLGTGIIISLIYQWL 854
QY 301 TLLLLAIVPIAIVAGVVMKMLSGQALKDKKEGAKIATEAIENTVTVSVLTQEKFE 360
Db 855 TLLLLAIVPIAIVAGVVMKMLSGQALKDKKEGAKIATEAIENTVTVSVLTQEKFE 914
QY 361 HMYAQSLQ 368
Db 915 NMYAQSLQ 922

RESULT 3
MDR3_MOUSE
ID MDR3_MOUSE STANDARD; PRT; 1276 AA.
AC P21447;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (P-glycoprotein 3) (MDR1A).
GN ABCB4 OR PGP3 OR PGP-3 OR MDR3 OR MDR1A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205845; PubMed=1969610;
RA Devault A., Gros P.;
RT "Two members of the mouse mdr gene family confer multidrug resistance
RL Mol. Cell. Biol. 10:1652-1663(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90287150; PubMed=1972547;
RA Hsu S.I.H., Cohen D., Kirschner L.S., Lothstein L., Hartstein M.,
RA Horwitz S.B.;
RT "Structural analysis of the mouse mdr1a (P-glycoprotein) promoter
RT reveals the basis for differential transcript heterogeneity in
RL Mol. Cell. Biol. 10:3596-3606(1990).
RN [3]
RP SEQUENCE OF 173-1276 FROM N.A.
RC STRAIN=BALE/C;
RX MEDLINE=89308614; PubMed=2473069;
RA Hsu S.I.H., Lothstein L., Horwitz S.B.;
RT "Differential overexpression of three mdr gene family members in
RT multidrug-resistant J774.2 mouse cells. Evidence that distinct P-
RT glycoprotein precursors are encoded by unique mdr genes.";
RL J. Biol. Chem. 264:12053-12062(1989).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
CC RELATED BUT DISTINCT CELLULAR GENES.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
-----
CC  EMBL: M30697; AAA39517.1; -
DR  EMBL: M33581; AAA39514.1; -
DR  EMBL: M33580; AAA39518.1; -
DR  EMBL: M2417; AAA03243.1; -
DR  PIR: A34175; DVMSIA.
DR  PIR: A34786; A34786.
DR  MGI: MGI:97570; Abcb4.
DR  InterPro: IPR003593; AAA.
DR  InterPro: IPR001140; ABC_transporter_tmern.
DR  InterPro: IPR003439; ABC_transporter.
DR  InterPro: IPR001687; ATP_GTP_A.
DR  Pfam: PF00664; ABC_membrane; 2.
DR  Pfam: PF00005; ABC_tran; 2.
DR  SMART: SM00382; AAA; 2.
DR  PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR  ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW  Multigene family.
FT  DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM 48 71
FT  TRANSMEM 117 136 POTENTIAL.
FT  TRANSMEM 188 205 POTENTIAL.
FT  TRANSMEM 212 232 POTENTIAL.
FT  TRANSMEM 298 318 POTENTIAL.
FT  TRANSMEM 327 346 POTENTIAL.
FT  TRANSMEM 347 707 CYTOPLASMIC (POTENTIAL).
FT  DOMAIN 708 728 POTENTIAL.
FT  TRANSMEM 753 773 POTENTIAL.
FT  TRANSMEM 829 848 POTENTIAL.
FT  TRANSMEM 853 880 POTENTIAL.
FT  TRANSMEM 942 961 POTENTIAL.
FT  TRANSMEM 964 984 POTENTIAL.
FT  DOMAIN 985 1276 CYTOPLASMIC (POTENTIAL).
FT  NP_BIND 423 430 ATP (BY SIMILARITY).
FT  NP_BIND 1066 1073 ATP (BY SIMILARITY).
FT  REPEAT 1 632
FT  REPEAT 633 1276
FT  CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CONFLICT 526 527 OL -> HV (IN REF. 2 AND 3).
FT  CONFLICT 939 939 F -> S (IN REF. 2 AND 3).
FT  CONFLICT 1036 1036 F -> V (IN REF. 2 AND 3).
SQ  SEQUENCE 1276 AA; 140754 MW; 75C1F3E1F58481 CRC64;

Query Match
Best Local Similarity 86.3%; Score 1584; DB 1; Length 1276;
Matches 317; Conservative 22; Mismatches 29; Indels 0; Gaps 0;

QY 1 TSALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGSGVVEQGNHDELMK 60
DB 554 TSALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGSGVVEQGNHDELMK 613
QY 61 EKGIFKLVMTQGRNEIDLENATGSKESDSEDALEMSPKDSGLKRSTRSIIHAPQG 120
DB 614 EKGIFKLVMTQGRNEIDLENATGSKESDSEDALEMSPKDSGLKRSTRSIIHAPQG 120
QY 121 QDRKLGTEDLNENPVVPSFWRLKLNSTWEPYVVGIFCAIINGGLQPAFISFRIIG 180
DB 674 QDRKLSTKEALDEDVPPASFWRLKLNSTWEPYVVGIFCAIINGGLQPAFISFRIIG 180
QY 181 IFTRDEDPETKROISNMFVLFVLGIISFTFFLOGFTFGKAGEILLRLRYMVFRLM 240
DB 734 VFTNGGPPETQRQNSLFLSLFLLIILGIIISFTFFLOGFTFGKAGEILLRLRYMVFRLM 240
QY 241 RQDVSFDDKNTTGALTFLRLANDAAQVKGAGISRLAVITONTANIGTGIISLIYGNOL 300
DB 794 RQDVSFDDKNTTGALTFLRLANDAAQVKGAGISRLAVITONTANIGTGIISLIYGNOL 300
QY 301 TLLLAIVPIIAIAGVVMKMLSGQALKDKKELEGAGKIATEAIEENFRTVWSLTQEKFE 360
DB 854 TLLLAIVPIIAIAGVVMKMLSGQALKDKKELEGAGKIATEAIEENFRTVWSLTQEKFE 360

QY 361 HMYAQSLSQ 368
DB 914 TMYAQSLSQ 921

RESULT 4
MDRI_RAT
ID MDRI_RAT STANDARD; PRT: 1277 AA.
AC P43245;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein 1).
GN ABCB1 OR PGY1 OR MDRI OR MDRI1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Silverman J.A., Raunio H., Gant T.W., Thorgeirsson S.S.;
RT "Cloning and characterization of a member of the rat multidrug
resistance (mdr) gene family.";
RL Gene 106:229-236(1991).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- PTM: SEVERAL PHOSPHORYLATED SERINE RESIDUES ARE PRESENT IN THE
LINKER DOMAIN.
CC -!- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
RELATED BUT DISTINCT CELLULAR GENES.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: M81855; -; NOT_ANNOTATED_CDS.
DR  InterPro: IPR003593; AAA.
DR  InterPro: IPR001140; ABC_transporter_tmern.
DR  InterPro: IPR003439; ABC_transporter.
DR  InterPro: IPR001687; ATP_GTP_A.
DR  Pfam: PF00664; ABC_membrane; 2.
DR  Pfam: PF00005; ABC_tran; 2.
DR  SMART: SM00382; AAA; 2.
DR  PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR  ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW  Multigene family.
FT  DOMAIN 1 47 CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM 48 71
FT  TRANSMEM 119 139 POTENTIAL.
FT  TRANSMEM 199 219 POTENTIAL.
FT  TRANSMEM 296 316 POTENTIAL.
FT  TRANSMEM 327 347 POTENTIAL.
FT  DOMAIN 348 709 CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM 710 730 POTENTIAL.
FT  TRANSMEM 755 775 POTENTIAL.
FT  TRANSMEM 832 852 POTENTIAL.
FT  TRANSMEM 854 874 POTENTIAL.
FT  TRANSMEM 937 957 POTENTIAL.
FT  TRANSMEM 968 988 POTENTIAL.
FT  DOMAIN 989 1277 CYTOPLASMIC (POTENTIAL).
FT  NP_BIND 426 433 ATP (BY SIMILARITY).
FT  NP_BIND 1070 1077 ATP (POTENTIAL).
FT  REPEAT 1 635
FT  REPEAT 636 1277
FT  CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```

SQ SEQUENCE 1277 AA; 141386 MW; 8AFDDDD619D2934C1 CRC64;
Query Match 77.3%; Score 1418.5; DB 1; Length 1277;
Best Local Similarity 74.3%; Pred. No. 4.5e-98;
Matches 275; Conservative 56; Mismatches 36; Indels 3; Gaps 2;

QY 1 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 60
DB 557 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 616
QY 61 EKGIIYFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSSSLIKRRSTRSHAPQG 120
DB 617 EKGIIYFKLVMTQTRGNEIDLENATGESKESDALEMSPKDSSSLIKRRSTRSHAPQG 675
QY 121 QDRKLGKEDLNENPPVSVFWRILKLNSTWPFYFVVGFFCALINGLOPAFISIRIIG 180
DB 676 QDRKLGKEDLNENPPVSVFWRILKLNSTWPFYFVVGFFCALINGLOPAFISIRIIG 735
QY 181 IPTREDDEPTKQISNMFSVLFLVGLIISITFFLQSGFTFGKAGEILTKRLRYMVFRSML 240
DB 736 VFSRDDHETKQISNMFSVLFLVGLIISITFFLQSGFTFGKAGEILTKRLRYMVFRSML 795
QY 241 RODVSWFDDKNTGALTRLANDAAQVKGAGISRLAVITONIANLGTGIIIS--LYGW 298
DB 796 RODVSWFDDKNTGALTRLANDAAQVKGAGISRLAVITONIANLGTGIIIS--LYGW 855
QY 299 QUTLLLLAIVPIIAAGVYEMKLSGQALKDKKEGAGKIATRAIENFRVVSILTQOK 358
DB 856 QUTLLLLAIVPIIAAGVYEMKLSGQALKDKKEGAGKIATRAIENFRVVSILTQOK 915
QY 359 FEHMYAQSLQ 368
DB 916 FEHMYAQSLQ 925

RESULT 5
MDRI_MOUSE STANDARD; PRT; 1276 AA.
ID MDRI_MOUSE AC P06795;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 1 (p-glycoprotein 1).
GN ABCB1 OR PGY1 OR PGY1-1 OR MDRI OR MDRI1B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87028229; PubMed=3768958;
RA Gros P., Croop J., Housman D.;
RT "Mammalian multidrug resistance gene: complete cDNA sequence
RT indicates strong homology to bacterial transport proteins.";
RN Cell 47:371-380(1986).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89367274; PubMed=2570420;
RA Raymond M., Gros P.;
RT "Mammalian multidrug-resistance gene: correlation of exon
RT organization with structural domains and duplication of an ancestral
RT gene.";
RN Proc. Natl. Acad. Sci. U.S.A. 86:6488-6492(1989).
[3]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=91042535; PubMed=2248681;
RA Raymond M., Gros P.;
RT "Cell-specific activity of cis-acting regulatory elements in the
RT promoter of the mouse multidrug resistance gene mdrl.";
RN Mol. Cell. Biol. 10:6036-6040(1990).
CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

```

Db 796 RQDISWFDHDKNSTGSLTRFLASDASSYKANGARLAVVQNVNGLGTGVILSLVYQWL 855
QY 301 TLLLLAIVPIIAAGVVKMISGQALKDKKELEGAGKIATEAIENFRTVSLTQEQKFE 360
Db 856 TLLLVVVIPLVLGGIEMKLLSGQALKDKKQLEISGKIATEAIENFRTVSLTQEQKFE 915
QY 361 HMYAQSLQ 368
Db 916 TMYAQSLQ 923

RESULT 6

MDR2_CRIGR STANDARD; PRT; 1276 AA.
AC F2L1449;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 2 (P-glycoprotein 2).
GN PGY2 OR PGP2.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92135896; PubMed=1685679;
RA Endicott J.A., Sarangi F., Ling V.;
RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
gene family";
RL DNA Seq. 2:89-101(1991).
[2]
RN SEQUENCE OF 622-1276 FROM N.A.
RX MEDLINE=88122132; PubMed=2893255;
RA Endicott J.A., Juranka P.F., Sarangi F., Gerlach J.H., Deuchars K.L.,
RA Ling V.;
RT "Simultaneous expression of two P-glycoprotein genes in
drug-sensitive Chinese hamster ovary cells.";
RL Mol. Cell. Biol. 7:4075-4081(1987).
CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT
CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
CANNOT.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M60041; AAA68884.1; -;
DR EMBL; M17896; AAA37007.1; -;
DR PIR; B27126; DVH32C.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmem.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 52 72 POTENTIAL.
FT TRANSMEM 119 139

FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT DOMAIN 346 708 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 709 729 POTENTIAL.
FT TRANSMEM 755 775 POTENTIAL.
FT TRANSMEM 831 851 POTENTIAL.
FT TRANSMEM 852 872 POTENTIAL.
FT TRANSMEM 935 955 POTENTIAL.
FT TRANSMEM 972 992 POTENTIAL.
FT DOMAIN 993 1276 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (POTENTIAL).
FT NP_BIND 1068 1075 ATP (POTENTIAL).
FT REPEAT 1 635
FT REPEAT 636 1276
SQ SEQUENCE 1276 AA; 141057 MW; 5096B1385628812D CRC64;

Query Match 74.4%; Score 1366.5; DB 1; Length 1276;
Best Local Similarity 72.8%; Pred. No. 3.4e-94;
Matches 268; Conservative 53; Mismatches 46; Indels 1; Gaps 1;

QY 1 TSALDTESEAVVQALDKARKGRITIVIAHRLSTVRNADVIAGFDGIVIVEKGNHDELMK 60
Db 557 TSALDTESEAVVQALDKAREGRITIVIAHRLSTVRNADVIAGFDGIVIVEQGNHEELMK 616
QY 61 EKGIVKLVMTQTRGNEIDLENATGESKSDALEMSPKDSLSLKRRTSRHSIHAPOG 120
Db 617 EKGIVKLVMTQTRGNEVELGSEADGSDSDTIASELITSEEPKPSV-RKSTCRSICGSD 675
QY 121 QDRKLGTKEDLNENVPVSWRILKLNSTENPYFVWGIFCAIINGLOPAPSIIFSRIIG 180
Db 676 QERRVSVKEAQDEDVPLVSFWGLKLNITENPYLVGVLCAVINGCMQPVFSIVFSGIIG 735
QY 181 IFTRDEDPETKROIKNMESVLFVLGIISFTTFEFGTFCGAGEILTCLRKYVWVFRSML 240
Db 736 VTRDDDDPKTKQONLFSFLFVGMGICFVYFFQGTFCGAGEILTCLRKYVWVFRSML 795
QY 241 RQDVSWFDDLNKTTGALTTRLANDAAQVKAIGSLRAVITONIANLGTGIIISLIYQWL 300
Db 796 RQDISWFDHDKNSTGSLTRFLASDAAVNVKGMASRLAGITQNVANLGTGIIISLIYQWL 855
QY 301 TLLLLAIVPIIAAGVVKMISGQALKDKKELEGAGKIATEAIENFRTVSLTQEQKFE 360
Db 856 TLLLVVVIPLVLGGIEMKLLSGQALKDKKQLEISGKIATEAIENFRTVSLTQEQKFE 915
QY 361 HMYAQSLQ 368
Db 916 NMYAQSLQ 923

RESULT 7

MDR3_HUMAN STANDARD; PRT; 1279 AA.
AC P21439;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (P-glycoprotein 3).
GN ABCB4 OR PGY3 OR MDR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89138016; PubMed=2906314;
RA van der Bliek A.M., Koolman P.M., Schneider C., Borst P.;
RT "Sequence of mdr3 cDNA encoding a human P-glycoprotein.";
RL Gene 71:401-411(1988).
RN [2]
RP SEQUENCE OF 856-1279 FROM N.A.
RX MEDLINE=88111519; PubMed=2892668;

RA van der Blik A.M., Baas F., ten Houte de Lange T., Koolman P.M.,
RA van der Velde-Koerts T., Borst P.;
RT "The human mdr3 gene encodes a novel P-glycoprotein homologue and
RT gives rise to alternatively spliced mRNAs in liver.";
RL EMBO J. 6:3325-3331(1987).
RN [3]
RP GENE STRUCTURE.
RX MEDLINE=91161629; PubMed=2002063;
RA Lincke C.R., Smit J.J.M., van der Velde-Koerts T., Borst P.;
RT "Structure of the human MDR3 gene and physical mapping of the human
RT MDR locus".;
RL J. Biol. Chem. 266:5303-5310(1991).
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS. HUMAN MDR3 IS NOT
CC CAPABLE OF CONFERRING DRUG RESISTANCE. MEDIATES THE TRANSLLOCATION
CC OF PHOSPHATIDYLCHOLINE ACROSS THE CANALICULAR MEMBRANE OF THE
CC HEPATOCYTE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- DISEASE: DEFECTS IN ABCB4 ARE A CAUSE OF PROGRESSIVE FAMILIAL
CC INTRAHEPATIC CHOLESTASIS TYPE III (PFIC), A FORM OF AUTOSOMAL
CC RECESSIVE LIVER DISORDERS, CHARACTERIZED BY EARLY ONSET OF
CC CHOLESTASIS THAT PROGRESSES TO CIRRHOSIS AND LIVER FAILURE BEFORE
CC ADULTHOOD.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M23234; AAA36207.1; -;
CC EMBL; X06181; CAA29547.1; -;
CC PIR; JS0051; DVH03.
CC PIR; A42213; A42213.
CC HSSP; P13569; 1NBD.
CC MIM; 171060; -;
CC MIM; 602347; -;
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001140; ABC_transporter_tmnm.
CC InterPro; IPR003439; ABC_transporter.
CC InterPro; IPR001687; ATP_GTP_A.
CC Pfam; PF00664; ABC_membrane; 2.
CC Pfam; PF00005; ABC_tran; 2.
CC SMART; SM00382; AAA; 2.
CC PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 192 211 POTENTIAL.
FT TRANSMEM 216 235 POTENTIAL.
FT TRANSMEM 301 320 POTENTIAL.
FT TRANSMEM 336 354 POTENTIAL.
FT DOMAIN 355 711 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 712 732 POTENTIAL.
FT TRANSMEM 756 776 POTENTIAL.
FT TRANSMEM 832 851 POTENTIAL.
FT TRANSMEM 854 873 POTENTIAL.
FT TRANSMEM 937 956 POTENTIAL.
FT TRANSMEM 976 993 POTENTIAL.
FT DOMAIN 994 1279 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 429 436 ATP (BY SIMILARITY).
FT NP_BIND 1069 1076 ATP (BY SIMILARITY).
FT REPEAT 1 640
FT REPEAT 641 1279
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1093 1093 V -> VFVDFGQ (IN REF. 2).
SQ SEQUENCE 1279 AA; 140682 MW; 3D58C9885C8D6087 CRC64;

Query Match 68.0%; Score 1248.5; DB 1; Length 1279;
Best Local Similarity 69.6%; Pred. No. 2.2e-85;
Matches 256; Conservative 42; Mismatches 65; Indels 5; Gaps 4;

QY 1 TSALDTESEAVVOVALDKARKGRITIVIAHRLSTVRNADVIAGFDGIVIVEKGNHDELMK 60
Db 560 TSALDTESEAEVQAALDKAREGRTIVIAHRLSTVRNADVIAGFDGIVIVEQSHSELMK 619
QY 61 EKIYFKLVMTQTRGNEIDLENATGESKSDALEMSPKDSGLIKRRSTRRSIHAPQG 120
Db 620 KEGVYFKLVNMTSGSQIOSEEF--ELNDEKAATRAPNWKSRFL-RHSTQKNLNSQM 676
QY 121 QDRKLGTKEK-LNENPPVSWFWILKLNSTENWYFVVGICAIINGLOPAPFISIRII 179
Db 677 CQKSLDVTGDLNANVPVSVFLKLNKTEWYFVVGIVCAITANGLOPAPFISIFSEII 736
QY 180 GIKTRDSDPTKQISNMFSVLFLVLGILSIFITFFLQGTGKAGELTKRLRYMVRFSM 239
Db 737 AIFGPGDD-AVKQKCNIFSLIFLGLISFTFFLQGTGKAGELTKRLRYMVRFSM 795
QY 240 LRQDVSWFDLKNVTGALTTRRLANDAAQVKGATGSLAVITONIANLGTGIIISLIYGM 299
Db 796 LRQDMSWFDHKNSTGALSTRLATAAQVQATGTRALIAQNIANLGTGIIISLIYGM 855
QY 300 LTLILLAIPIIAIAGVVEKMLSGOALKDKKELEGAGKATIAENFRIVVSLTQEQKF 359
Db 856 LTLILLAVPIIAVSGIVEMKLLAGNAKDKKELEGAGKATIAENFRIVVSLTQEQKF 915
QY 360 EHYMAQSL 367
Db 916 ESMYVEKL 923

RESULT 8

MDR2_MOUSE
ID MDR2_MOUSE STANDARD; PRT; 1276 AA.
AC P21440;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 2 (P-glycoprotein 2).
GN PGY2 OR PGY-2 OR MDR2.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88302195; PubMed=3405218;
RA Gros P., Raymond M., Bell J., Housman D.;
RT "Cloning and characterization of a second member of the mouse mdr
RT gene family.";
RL Mol. Cell. Biol. 8:2770-2778(1988).
RN [2]
RP SEQUENCE OF 1-23 FROM N.A.
RC STRAIN=BA1B/C;
RA Kirschner L.S., Horwitz S.B.;
RL Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS. MOUSE MDR2 IS
CC NOT CAPABLE OF CONFERRING DRUG RESISTANCE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- MISCELLANEOUS: IN MOUSE THE MDR GENE FAMILY INCLUDES THREE OR MORE
CC RELATED BUT DISTINCT CELLULAR GENES.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC

or send an email to license@isb-sib.ch).

```
CC -----
CC EMBL; J03398; AAA39516.1; -.
DR EMBL; M74151; AAA39515.1; -.
DR PIR; A30409; DVMS2.
DR HSSP; P13569; INBD.
DR MGD; MGI:97569; Pgy2.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transporter_tmern.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family.
FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 52 75 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT DOMAIN 346 707 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 753 773 POTENTIAL.
FT TRANSMEM 829 849 POTENTIAL.
FT TRANSMEM 850 870 POTENTIAL.
FT TRANSMEM 933 953 POTENTIAL.
FT TRANSMEM 970 990 POTENTIAL.
FT DOMAIN 991 1276 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 426 433 ATP (BY SIMILARITY).
FT NP_BIND 1066 1073 ATP (BY SIMILARITY).
FT REPEAT 1 635
FT REPEAT 636 1276
FT SEQUENCE 1276 AA; 140332 MW; A6C38DCD2C118EEF CRC64;

Query Match 66.5%; Score 1220.5; DB 1; Length 1276;
Best Local Similarity 67.6%; Pred. No. 2.6e-83;
Matches 248; Conservative 49; Mismatches 67; Indels 3; Gaps 3;

QY 1 TSALDTESEAVVOALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 60
DB 557 TSALDTESEAEVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVVEKGNHDELMK 616
QY 61 EKIYKLVMTQTRGNEIDLENATGESKESDALEMSPKDSGLIKRSTRSHAPQG 120
DB 617 KEGYIFRLVNMOTSGSQLSEFEVELEDEKAAGVAP-NGWKARIFRNSTKSLKSPH- 674
QY 121 QDRKLGTKEDLNENVPVPSFWRILKLNSTEWPPYVVGIFCAINGLOPAFSIIFRITG 180
DB 675 QNRLEDTEENLDANVPVPSFLKVLKLNKTEWPPYVVGTVCAITANGALQPAFSIILSEMA 734
QY 181 IFRDEDEPTKQISNMFSVFLVGLIISFIFLPGFTFGKAGEILLTKRLRYWVPSML 240
DB 735 IFPGPGDD-AVKQKQCNMFSVFLVGLIISFIFLPGFTFGKAGEILLTKRLRYWVPSML 793
QY 241 RDVSWFDDLNKNTGALTTRLANDAAQVKAIGSRSLAVITQNTANLGTGIIISLYGWQL 300
DB 794 RQMSWFDHDKNSTGALSTRLATDAQVQAGTGTALTAQNTANLGTGIIISLYGWQL 853
QY 301 TLLLAITVPIAIAVGVEMKMLSGQALKDKKLEAGKATEAIEAENFRVWVLSIQPKFE 360
DB 854 TLLLSVVPFIAVAGIVEMKMLAGNAKDKKLEAGKATEAIEAENFRVWVLSIQPKFE 913
QY 361 HMYAQLS 367
DB 914 SMYVEKL 920

.
RESULT 9
MDR2_RAT
```



```
QY 121 QDRKLGTK-EDLNENVPVSEWRILKLNSTWEPYFVVGIFCAIINGGLQPAFISIRII 179
Db 676 HONRLDVETNELDANVPVSKVLRLNKTEWYFVVGTLCAIANGALQPAFISILSEMI 735
QY 180 GIFTRDEDPETKROISNMFSLVFLVGLIISIFITFFLOGFTFGKAGEILTKRLRYWVFRSM 239
Db 736 AIFGPGDD-TVKQCKNMFSLVFLGLGVHVSFTFFLOGFTFGKAGEILTKRLRSMAPKAM 794
QY 240 LRQDVSWFDDLLKNTTGALTTRLANDAAQVKGAGISRLAVITONTANLGTGIISLIYQWQ 299
Db 795 LRQDSWFDDHKSTGALSTRLATDAAQVQAGTGLTRALIAQNTANLGTGIISLIYQWQ 854
QY 300 LTLILLAIPIIAIAGVVMKMLSGQALKDKKEGAGKATEAIEFNRTVWSLTQQRKF 359
Db 855 LTLILLSVVPPIAVAGIVEMKMLAGNAKRDKKEMEAGKATEAIEFNRTVWSLTQQRKF 914
QY 360 EHYMAQSL 367
Db 915 ESMYVEKL 922

RESULT 10
MDR3_CRIGR STANDARD; PRT; 1281 AA.
ID MDR3_CRIGR STANDARD; PRT; 1321 AA.
AC P23174;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (P-glycoprotein 3).
GN PGY3 OR PGP3.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92135896; PubMed=1685679;
RA Endicott J.A., Sarangi F., Ling V.;
RT *Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
RL gene family.
RL DNA Seq. 2:89-101(1991).
CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT
CC CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
CC CANNOT.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M60042; AAA68895.1; --
CC HSP; P13569; INBO.
CC InterPro: IPR003593; AAA.
CC InterPro: IPR001140; ABC_transporter_unem.
CC InterPro: IPR003439; ABC_transporter.
CC InterPro: IPR001687; ATP_GTP_A.
CC Pfam: PF00664; ABC_membrane; 2.
CC Pfam: PF00005; ABC_tran; 2.
CC SMART; SM00382; AAA; 2.
CC PROSITE; PS00211; ABC_TRANSPORTER; 1.
CC ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
CC Multigene family.
CC DOMAIN 1 57 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 58 78 POTENTIAL.
```

```
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 299 319 POTENTIAL.
FT TRANSMEM 328 348 POTENTIAL.
FT DOMAIN 349 712 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 713 733 POTENTIAL.
FT TRANSMEM 758 778 POTENTIAL.
FT TRANSMEM 834 854 POTENTIAL.
FT TRANSMEM 855 875 POTENTIAL.
FT TRANSMEM 938 958 POTENTIAL.
FT TRANSMEM 975 995 POTENTIAL.
FT DOMAIN 996 1281 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 429 436 ATP (POTENTIAL).
FT NP_BIND 1071 1078 ATP (POTENTIAL).
SQ SEQUENCE 1281 AA; 140866 MW; 2203EF61EBB29602 CRC64;

Query Match 65.6%; Score 1204.5; DB 1; Length 1281;
Best Local Similarity 66.6%; Pred. No. 4.2e-82;
Matches 245; Conservative 48; Mismatches 72; Indels 3; Gaps 3;

QY 1 TSALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGCVIVEKGNHDELMK 60
Db 560 TSALDTESEAEVQALDKARKGRTTIVIAHRLSTVRNADVIAGFDGCVIVEQGSHELMQ 619
QY 61 EKIYFKLVMTQRGNEIDLENATGESKSEDALESPKDSGLIKRSTRSRSHAPQG 120
Db 620 KESVFKLVNMTSGSILSQEFSELSDEKADGWP-NGWKSHIFRNSTRKSKSRSA 678
QY 121 QDRKLGTKED-LNENVPVSVFWRILKLNSTWEPYFVVGIFCAIINGGLQPAFISIRII 179
Db 679 HHRLDVDADDELDAVNPVPSFLKVLKLNTEWYFVVGTVCAIVNGALQPAISILSEMI 738
QY 180 GIFTRDEDPETKROISNMFSLVFLVGLIISIFITFFLOGFTFGKAGEILTKRLRYWVFRSM 239
Db 739 AIFGPGDD-AVKQCKNLSLVFLGLGVHVSFTFFLOGFTFGKAGEILTKRLRSMAPKAM 797
QY 240 LRQDVSWFDDLLKNTTGALTTRLANDAAQVKGAGISRLAVITONTANLGTGIISLIYQWQ 299
Db 798 LRQDSWFDDYKNSGALSTRLATDAAQVQAGTGLTRALIAQNTANLGTGIISLIYQWQ 857
QY 300 LTLILLAIPIIAIAGVVMKMLSGQALKDKKEGAGKATEAIEFNRTVWSLTQQRKF 359
Db 858 LTLILLSVVPPIAVAGIVEMKMLAGNAKRDKKEMEAGKATEAIEFNRTVWSLTQQRKF 917
QY 360 EHYMAQSL 367
Db 918 ESMYVEKL 925

RESULT 11
ABILL_RABIT
ID ABILL_RABIT STANDARD; PRT; 1321 AA.
AC Q9NOV3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11)
DE (Sister of P-glycoprotein).
OS ABCB11 OR BSEP OR SPGP.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
RA Balasubramanian N.V., Suchy F.J., Ananthanarayanan M.;
RT "Molecular cloning and characterization of rabbit liver bile salt
RT export pump (Bsep/spgp).";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
CC INTO THE CANALICULUS OF HEPATOCYTES.
```

```
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC -|- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
CC MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
CC SITU.
CC -|- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF249879; AAF65552.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001887; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transmembrane; Transport.
FT DOMAIN 1 62
FT TRANSMEM 63 83
FT DOMAIN 84 147
FT TRANSMEM 148 168
FT DOMAIN 169 215
FT TRANSMEM 216 236
FT DOMAIN 237 240
FT TRANSMEM 241 261
FT DOMAIN 262 319
FT TRANSMEM 320 340
FT DOMAIN 341 353
FT TRANSMEM 354 374
FT DOMAIN 375 755
FT TRANSMEM 756 776
FT DOMAIN 777 794
FT TRANSMEM 795 815
FT DOMAIN 816 869
FT TRANSMEM 870 890
FT DOMAIN 891 911
FT TRANSMEM 912 979
FT DOMAIN 980 1000
FT TRANSMEM 1001 1011
FT DOMAIN 1012 1032
FT TRANSMEM 1033 1321
FT NP_BIND 455 462
FT NP_BIND 1113 1120
FT CARBOHYD 109 109
FT CARBOHYD 116 116
FT CARBOHYD 122 122
FT CARBOHYD 125 125
FT SEQUENCE 1321 AA; 146376 MW; 457539FCD6D717A2 CRC64;
Query Match 43.8%; Score 803.5; DB 1; Length 1321;
Best Local Similarity 42.0%; Pred. No. 3.5e-52;
Matches 163; Conservative 84; Mismatches 116; Indels 25; Gaps 4;
QY 1 TSALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGVTVEKGNHDELMK 60
DB 586 TSALDTESEAVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGVTVEKGNHDELMK 645
QY 61 EKGIVFKLVTMOTRGNEIDLE-----NATGESKSEDALEMSPKDSGLIKRRS----- 110
DB 646 RKGIVFALVTLSQNRQNGDQEEENKATEDDIPKFTSRGNQDSLRSLRQSRKLSY 705
QY 111 -----TRRSIHAPQODRKLGKTEDLNENPPVSVFWILLKLNSTENWFVVGIFC 160
```

```
Db 706 LAHEPPMAVEDHKSTHEEDRKDLPAQED-----IEPASVRRIMKLNAPWPMILGSMG 761
QY 161 AILINGLOPAFAISRIIGIFTRDQDPETKROISNMFSVLFLVLGLIISFIFFLQGF 220
Db 762 AAVNGAVTPYAFLLFOILGTFSLDPKEQORSI-NGICLLFTVLCVFFTFQLOGYTF 820
QY 221 GKAGEILTKRLRYMVFRLSRMLRQDVSWMFDDLNKNTGALTTRLANDAAQVKAIGSLAVIT 280
Db 821 AKSGELLTKRLRFGFRAMLGQDQIGWFDLRLNSPGALTTRLATDASQVQATGSGQIGMV 880
QY 281 QNTANLGTGIIISLIYQWQLTLLLLAIVPIATAGVVMKMLSGQALKKKELEGAGKIA 340
Db 881 NSETNVTVMATIAFLFSWKLTLGIVCFPPFLALSGALQTKMLTGFASRDKQALEKAGQIT 940
QY 341 TEAENFRTVSLTQEQKPEHMYAQSLQ 368
Db 941 SEALSNIPTVAGIGKRRKFTFEAELE 968
RESULT 12
AB11_MOUSE STANDARD; PRT; 1321 AA.
ID AB11_MOUSE Q9QZE8;
AC Q9QY30; Q9QZE8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11)
DE (Sister of P-glycoprotein).
GN ABCB11 OR BSEP OR SPGP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=20076398; PubMed=10607905;
RA Green R.M., Hoda F., Ward K.L.;
RT "Molecular cloning and characterization of the murine bile salt export
pump."
RL Gene 241:117-123(2000).
RN [2]
RP SEQUENCE OF 463-635 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RA Salkar R., Suchy F.J., Ananthanarayanan M.;
RT "Molecular cloning of mouse liver bile salt export pump (bsep).";
Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
INTO THE CANALICULUS OF HEPATOCYTES.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC -|- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
SITU.
CC -|- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
ATP BINDING CASSETTE (ABC) DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF133903; AAF14372.1; -.
DR EMBL; AF186585; AAF56419.1; -.
DR MGI; MGI:1351619; Abcb11.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
```

```
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transmembrane; Cysteine.
FT DOMAIN 1 62
FT TRANSMEM 63 83
FT DOMAIN 84 147
FT TRANSMEM 148 168
FT TRANSMEM 169 215
FT TRANSMEM 216 236
FT TRANSMEM 237 240
FT TRANSMEM 241 261
FT TRANSMEM 262 319
FT TRANSMEM 320 340
FT TRANSMEM 341 353
FT TRANSMEM 354 374
FT TRANSMEM 375 755
FT TRANSMEM 756 776
FT TRANSMEM 777 794
FT TRANSMEM 795 815
FT TRANSMEM 816 869
FT TRANSMEM 870 890
FT TRANSMEM 891 911
FT TRANSMEM 912 979
FT TRANSMEM 980 1000
FT TRANSMEM 1001 1011
FT TRANSMEM 1012 1032
FT TRANSMEM 1033 1321
FT NP_BIND 455 462
FT NP_BIND 1113 1120
FT CARBOHYD 109 109
FT CARBOHYD 116 116
FT CARBOHYD 122 122
FT CARBOHYD 125 125
FT CARBOHYD 125 125
FT CONFLICT 481 481
FT CONFLICT 633 633
SQ SEQUENCE 1321 AA; 146675 MW; 15B5EBF175D32967 CRC64;

Query Match 43.3%; Score 794.5; DB 1; Length 1321;
Best Local Similarity 43.0%; Pred. No. 1.7e-51;
Matches 166; Conservative 77; Mismatches 122; Indels 21; Gaps 5;

QY 1 TSALTESEAVQVVALDKARKGRITIVIAHRLSTVRNADVTAGDDGVIVEKGNHDELMK 60
Db 586 TSALNESEAKVQGNALNKIQHGTIIISVAHRLSTVRNADVTAGDDGVIVEKGNHDELMK 645
QY 61 EKGIVFKLVMTQTRNEIDLENA-TGESKSESDALEM-----SPKDGSSSLIKRRSTR- 113
Db 646 RKGIVFKLVMTQTRNEIDLENA-TGESKSESDALEM-----SPKDGSSSLIKRRSTR- 113
QY 114 -----SIHAPOQDRKLGTCKEDLNENPVVSVFWRILKLNSTWPFVYVGFICAI 162
Db 706 SHLSHEPPLATGDKSSYEDRK--DNDVLVEVEPAPVRRILKYNISEWPYILVGCALCA 763
QY 163 INGGLOPAFISFIISRIIGITRDEDPETKROISNMFSLVFLVGLIISFITFLQGFTEGK 222
Db 764 INGAVTPTIYSLFSLQILKFTSLVDKEQORSEIYSN-CLFFVILGCVSLFTQFQGNFAK 822
QY 223 AGEILTKRLRYWFRSMRLQRQVSDFDLLKNTTGALTTLRLANDAAQVKGAGSLRAVITON 282
Db 823 SGELLTKRLRFKFGKLNKQDIDGFDLDDLNKPNGLVITTLRLANDAAQVKGAGSLRAVITON 282
QY 283 TANLGTGIISLIYQWLTLLILATVPIIAIAGVVENKMLSGQALKKKKEGAGKIATE 342
Db 883 FTNIEFVAVLIAFLFNWKLISLVISFFPLALSGAVQTKMLTGFASQDKKLEILKAGQITNE 942
QY 343 AIENRPTVSVSTQEQKFHMYAQSLQ 368
Db 943 ALSNIRTVAGIGVEGRFKAFKFEVELE 968
```

```
RESULT 13
AB11_HUMAN
ID AB11_HUMAN STANDARD; PRT; 1321 AA.
AC O95342; Q9UNB2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bile salt export pump (ATP-binding cassette, sub-family B, member 11).
GN ABCB11 or BSEP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT PFIC2 GLY-297.
RX MEDLINE=99021377; PubMed=9806540;
RA Strautnieks S.S., Bull L.N., Knisely A.S., Kocoshis S.A., Dahl N.,
RA Arnell H., Sokal E., Dahan K., Childs S., Ling V., Tanner M.S.,
RA Kagalwalla A.F., Nemeth A., Pawlowska J., Baker A., Mieli-Vergani G.,
RA Feiner N.B., Gardiner R.M., Thompson R.J.;
RT "A gene encoding a liver-specific ABC transporter is mutated in
progressive familial intrahepatic cholestasis.";
RL Nat. Genet. 20:233-238(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Mol O., Hoolveld G.J.E.J., Jansen P.L.M., Muller M.;
RT "Cellular localization and functional characterization of the human
bile salt export pump (BSEP).";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
INTO THE CANALICULUS OF HEPATOCYTES.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR
MICROVILLI AND TO SUBCANALICULAR VESICLES OF THE HEPATOCYTES BY IN
SITU (BY SIMILARITY).
CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
ATP BINDING CASSETTE (ABC) DOMAIN.
CC -1- DISEASE: DEFECTS IN ABCB11 ARE THE CAUSE OF PROGRESSIVE FAMILIAL
INTRAHEPATIC CHOLESTASIS 2 (PFIC2), AN INHERITED LIVER DISEASE OF
CHILDHOOD. PFIC2 IS CHARACTERIZED BY CHOLESTASIS AND NORMAL SERUM
GAMMA-GUTAMYLTRANSFERASE ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF091582; AAC77455.1; -.
CC EMBL; AF136523; AAD28285.1; -.
CC MIM; 603201; -.
CC MIM; 601847; -.
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001140; ABC_transporter_tmem.
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR001687; ATP_GTP_A.
CC Pfam; PF00664; ABC_membrane; 2.
CC Pfam; PF00005; ABC_tran; 2.
CC SMART; SM00382; AAA; 2.
CC PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transmembrane; Transport; Disease mutation.
FT DOMAIN 1 62
FT TRANSMEM 63 83
FT TRANSMEM 84 147
FT TRANSMEM 148 168
FT TRANSMEM 169 215
FT TRANSMEM 216 236
```



```
QY 361 HMYAQSILQ 368
Db 859 YMYAQSILQ 866

RESULT 4
Q60502
ID Q60502 PRELIMINARY; PRT; 1169 AA.
AC Q60502
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE P-GLYCOPROTEIN.
GN GGP-1.
OS Cricetus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=91154265; PubMed=1671863;
RA Devine S.E., Hussain A., Davide J.P., Melera P.W.;
RT "Full length and alternatively spliced ggp-1 transcripts in multidrug
resistant Chinese hamster lung cells.";
RL J. Biol. Chem. 266:4545-4555(1991).
CC 1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; M59254; AAA37005.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tm.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1169 AA; 128939 MW; 72E25B7CE29DC185 CRC64;

Query Match 86.5%; Score 1589; DB 11; Length 1169;
Best Local Similarity 85.1%; Pred. No. 1.7e-125;
Matches 313; Conservative 29; Mismatches 26; Indels 0; Gaps 0;

QY 1 TSALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVIAGDGGVIVEKGNHDELMK 60
Db 448 TSALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVIAGDGGVIVEKGNHDELMR 507

QY 61 EKGIFKLVMTQTRGNEIDLENATGESKSDALEMSPKDSGLIKRRSTRRSIHAPOG 120
Db 508 EKGIFKLVMTQTAGNEIENGEVGEKNEIDNLDMSKDSASSLIIRSTRRSIRPHD 567

QY 121 QDRKLGKEDLNENPVPSFWRILKLNSTWEPYFVVGIFCAINGLQPAFSIIFSRIG 180
Db 568 QDRKLTKEALDEDPPIPSFWRILKLNSTWEPYFVVGIFCAIVNGALQPAFSIIFSKVG 627

QY 181 IFTRDEDPETKROI SNMFSVLFLVLGIISFIFTFLOGFTFGKAGEILTTLRLYMWVFRSML 240
Db 628 VFTRTDDETKRHSNLSFLFLILGIVISFIFTFLOGFTFGKAGEILTTLRLYMWVFRSML 687

QY 241 RQDVSWFDDLNKNTTGALTTRLANDAAQVGAIGSRSLAVITONIANLGTGIIISLIYGMOL 300
Db 688 RQDVSWFNDPNKNTTGALTTRLANDAQVKGATGARLAVITONIANLGTGIIISLIYGMOL 747

QY 301 TLLLLAIVPIIAIAGVVMKMLSGQALKDKKELEGAGKIATEAENFRVVSILTQEQKFE 360
Db 748 TLLLLAIVPIIAIAGVVMKMLSGQALKDKKELEGAGKIATEAENFRVVSILTQEQKFE 807

QY 361 HMYAQSILQ 368
Db 808 HMYAQSILQ 815
```

```
RESULT 5
O02793
ID O02793 PRELIMINARY; PRT; 1285 AA.
AC O02793;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN-1.
GN MOR1.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Longley M., Crawford A.M.;
RT "Ovine mdrl gene.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; U78609; AAB58489.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tm.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1285 AA; 142020 MW; 90153E617C44856F CRC64;

Query Match 86.5%; Score 1589; DB 6; Length 1285;
Best Local Similarity 85.1%; Pred. No. 2e-125;
Matches 313; Conservative 33; Mismatches 22; Indels 0; Gaps 0;

QY 1 TSALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVIAGDGGVIVEKGNHDELMK 60
Db 563 TSALDTESEAVVQALDKARKGRTTIVIAHRLSTVRNADVIAGDGGVIVEKGNHDELMG 622

QY 61 EKGIFKLVMTQTRGNEIDLENATGESKSDALEMSPKDSGLIKRRSTRRSIHAPOG 120
Db 623 KRGIFKLVMTQTRGNEIENGEVGEKNEIDNLDMSKDSASSLIIRSTRRSIRGSQS 682

QY 121 QDRKLGKEDLNENPVPSFWRILKLNSTWEPYFVVGIFCAINGLQPAFSIIFSRIG 180
Db 683 QDRKLTSTEDLDSVPSFWRILKLNSTWEPYFVVGIFCAINGALQPAFSIIFSRIG 742

QY 181 IFTRDEDPETKROI SNMFSVLFLVLGIISFIFTFLOGFTFGKAGEILTTLRLYMWVFRSML 240
Db 743 IFTRNNDDETKRHSNLSFLFLILGIVISFIFTFLOGFTFGKAGEILTTLRLYMWVFRSML 802

QY 241 RQDVSWFDDLNKNTTGALTTRLANDAAQVGAIGSRSLAVITONIANLGTGIIISLIYGMOL 300
Db 803 RQDVSWFDDPNKNTTGALTTRLANDAAQVGAIGSRSLAVITONIANLGTGIIISLIYGMOL 862

QY 301 TLLLLAIVPIIAIAGVVMKMLSGQALKDKKELEGAGKIATEAENFRVVSILTQEQKFE 360
Db 863 TLLLLAIVPIIAIAGVVMKMLSGQALKDKKELEGAGKIATEAENFRVVSILTQEQKFE 922

QY 361 HMYAQSILQ 368
Db 923 YMYAQSILQ 930

RESULT 6
Q9JK64
ID Q9JK64 PRELIMINARY; PRT; 1272 AA.
AC Q9JK64;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
```

```
DE MULTIDRUG RESISTANCE PROTEIN 1A.
GN PCY1
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=LIVER;
RA Hooiveld G.J.E.J., Wilms J.W.J., Hagenbuch B., Jansen P.L.M.,
RA Meijer D.K.F., Muller M.;
RT "Cloning and functional characterization of the rat multidrug
RT resistance protein Mdr1a.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF257746; AAF69007.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane_2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1272 AA; 140326 MW; B5C1DF89DFC986C5 CRC64;

Query Match 85.8%; Score 1575; DB 11; Length 1272;
Best Local Similarity 84.8%; Pred. No. 3e-124;
Matches 312; Conservative 28; Mismatches 28; Indels 0; Gaps 0;

Qy 1 TSALDTESEAVVQALDKARKGRITIVIAHRLSTVRNADVIAGFDGQVIVEQGNHDELMK 60
Db 550 TSALDTESEAVVQALDKAREGTTIVIAHRLSTVRNADVIAGFDGQVIVEQGNHDELMR 609
Qy 61 EKGIFKLVMTQNRGNEIDLENATGESKSDALEMSPKDSGSLIKRRSTRRSIHAPQG 120
Db 610 EKGIFKLVMTQNRGNEIDLENATGESKSDGIDNVDMSSKDSGSLIKRRSTRRSIRPHD 669
Qy 121 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFYVGVFCALINGGLQPAFSIIFSRIG 180
Db 670 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFYVGVFCALINGGLQPAFSIIFSKVVG 729
Qy 181 IFTREDEPTKROISNMFSVLFLVLGIISITFFLQGTFTGKAGEILTTLRLRYMVFRSML 240
Db 730 VFTKNDTPEIQRQNSLFLVLGIISITFFLQGTFTGKAGEILTTLRLRYMVFRSML 789
Qy 241 RQDVSFDDLKNTGALTTRLANDAAQVKAIGSLAVITONTIANLGTGIIISLIYGWQL 300
Db 790 RQDISWFDPKNTGALTTRLANDAAQVKAIGSLAVITONTIANLGTGIIISLIYGWQL 849
Qy 301 TLLLLAIPIIAIAGVVEKMLSGQALKKKELEGAGKIAEIAENFRTVVSILTREQKFE 360
Db 850 TLLLLAIPIIAIAGVVEKMLSGQALKKKELEGAGKIAEIAENFRTVVSILTREQKFE 909
Qy 361 HMYAQSLQ 368
Db 910 TMYAQSLQ 917

RESULT 8
O93437 PRELIMINARY; PRT; 1288 AA.
ID O93437;
AC O93437;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE ABC TRANSPORTER PROTEIN.
GN CMDRI.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
EX MEDLINE=99209805; PubMed=10195430;
RA Edelman H.M.L., Duchek P., Rosenthal F.E., Foeger N., Glackin C.,
RA Kane S.E., Kuchler K.;
RT "Cmrl, a chicken P-glycoprotein, confers multidrug resistance and
RT interacts with Estradiol.";
RL Biol. Chem. 380:231-241(1999).
DR EMBL; AJ009799; CAA08835.1; -.
DR HSP; P13569; 1NED.
FT CHAIN.
FT SEQUENCE 1288 AA; 141917 MW; CB258A5F2826DB6C CRC64;

Query Match 65.3%; Score 1198; DB 13; Length 1288;
Best Local Similarity 65.7%; Pred. No. 2.3e-92;
Matches 249; Conservative 54; Mismatches 52; Indels 24; Gaps 9;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Yamazaki M., Leake B.F., Kim R.B.;
RA "Molecular Cloning of Rat Mdr1a cDNA.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF286167; AAK83023.1; -.
SQ SEQUENCE 1272 AA; 140343 MW; D9A17457362DA0ED CRC64;

Query Match 85.3%; Score 1566; DB 11; Length 1272;
Best Local Similarity 84.2%; Pred. No. 1.7e-123;
Matches 310; Conservative 28; Mismatches 30; Indels 0; Gaps 0;

Qy 1 TSALDTESEAVVQALDKARKGRITIVIAHRLSTVRNADVIAGFDGQVIVEQGNHDELMK 60
Db 550 TSALDTESEAVVQALDKAREGTTIVIAHRLSTVRNADVIAGFDGQVIVEQGNHDELMR 609
Qy 61 EKGIFKLVMTQNRGNEIDLENATGESKSDALEMSPKDSGSLIKRRSTRRSIHAPQG 120
Db 610 EKGIFKLVMTQNRGNEIDLENATGESKSDGIDNVDMSSKDSGSLIKRRSTRRSIRPHD 669
Qy 121 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFYVGVFCALINGGLQPAFSIIFSRIG 180
Db 670 QDRKLGTKEDLNENPVPSFWRILKLNSTWPFYVGVFCALINGGLQPAFSIIFSKVVG 729
Qy 181 IFTREDEPTKROISNMFSVLFLVLGIISITFFLQGTFTGKAGEILTTLRLRYMVFRSML 240
Db 730 VFTKNDTPEIQRQNSLFLVLGIISITFFLQGTFTGKAGEILTTLRLRYMVFRSML 789
Qy 241 RQDVSFDDLKNTGALTTRLANDAAQVKAIGSLAVITONTIANLGTGIIISLIYGWQL 300
Db 790 RQDISWFDPKNTGALTTRLANDAAQVKAIGSLAVITONTIANLGTGIIISLIYGWQL 849
Qy 301 TLLLLAIPIIAIAGVVEKMLSGQALKKKELEGAGKIAEIAENFRTVVSILTREQKFE 360
Db 850 TLLLLAIPIIAIAGVVEKMLSGQALKKKELEGAGKIAEIAENFRTVVSILTREQKFE 909
Qy 361 HMYAQSLQ 368
Db 910 TMYAQSLQ 917

RESULT 8
O93437 PRELIMINARY; PRT; 1288 AA.
ID O93437;
AC O93437;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE ABC TRANSPORTER PROTEIN.
GN CMDRI.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
EX MEDLINE=99209805; PubMed=10195430;
RA Edelman H.M.L., Duchek P., Rosenthal F.E., Foeger N., Glackin C.,
RA Kane S.E., Kuchler K.;
RT "Cmrl, a chicken P-glycoprotein, confers multidrug resistance and
RT interacts with Estradiol.";
RL Biol. Chem. 380:231-241(1999).
DR EMBL; AJ009799; CAA08835.1; -.
DR HSP; P13569; 1NED.
FT CHAIN.
FT SEQUENCE 1288 AA; 141917 MW; CB258A5F2826DB6C CRC64;

Query Match 65.3%; Score 1198; DB 13; Length 1288;
Best Local Similarity 65.7%; Pred. No. 2.3e-92;
Matches 249; Conservative 54; Mismatches 52; Indels 24; Gaps 9;
```

```
QY 1 TSALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVIVEKGNHDELMK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 566 TSALDTESEVVOAALDKAREGRTTVVAHRLSTVRNADVIAGFDGIVIVEKGNHDELMK 625
QY 61 EKGIIYFLVMTQTRGNEIDLENATGESKESDALEMSPKDSGS-----SL---IKRST 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 626 KGGIYKLVNMQT---IETDPSE-KSEN---AVSKRSGSGLNDESLKKELRRGST 677
QY 112 RRSIH---APOQDRKLGCTKEDLNENPVVPSFWRLTKLNSTWEPYFVVGIFCALINGLQ 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 678 RSMKKPCEPNDTEK-GSSPD---EELPPVFLKMLKLNKNEWPIYFVAGTFCALVNGALQ 734
QY 169 PAFSIIISRIIGITFRDEDPETKQISNMFSVLFLVGLIISITFFLQGGFTFGKAGEILT 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 735 PAFSVIFSEIIIGITS-ETDQVLRKSNLSLLFLALGLIISFFTFVQGFAGKAGEILT 793
QY 229 KRLRWVFRMLRDQSVFDDLNKNTGALTTRRLANDAAQVKGATGSRILAVITONIANLGT 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 794 MKLRFMAFKAMLRQDMAFWDDPKNSTGALTTRRLANDASQVKGATGVRLLALIAQIANLGT 853
QY 289 GIISLIYGVWLTLLLLAIVPIIAIAGVVENKMLSGOALKDKKELEGAGKIATEAIENR 348
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 854 GIISLIYGVWLTLLLLAIVPIIAIAGVMIEMKLAGHAKKDKIELEAAGKIATEAIENR 913
QY 349 TVVSLTQDQKFEHMYAOSL 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 914 TVASLTREKPELMVGEHL 932

RESULT 9
Q91586 PRELIMINARY; PRT; 1287 AA.
AC Q91586;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN.
GN XEMDR.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95322451; PubMed=7599185;
RA Castillo G., Shen H.J., Horwitz S.B.;
RT "A homologue of the mammalian multidrug resistance gene (mdr) is functionally expressed in the intestine of Xenopus laevis.";
RL Blochim. Biophys. Acta 1262:113-123(1995).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; U17608; AAA75000.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR ATP-binding; Transport.
SQ SEQUENCE 1287 AA; 141505 MW; 06E95073C5771415 CRC64;

Query Match 63.0%; Score 1156; DB 13; Length 1287;
Best Local Similarity 62.8%; Pred. No. 8.3e-89;
Matches 236; Conservative 61; Mismatches 61; Indels 18; Gaps 6;

QY 1 TSALDTESEAVVOVALDKARKGRTTIVIAHRLSTVRNADVIAGFDGIVIVEKGNHDELMK 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 568 TSALDTESEAVQALDKAREGRTTIVVAHRLSTVRNADVIAGFDGIVIVEQSHHELM 627
QY 61 EKGIIYFLVMTQMT---RGNEIDLENATGESKESDALEMSPKDSG---SSLIKRSTRESI 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 KGGVTHLVTMTQFKADE-----GEEDNLSAGEKSPHNNVIESPLLRKSTRESS 237
QY 116 HA-----PQGDRLKGTKEDLNENPVVPSFWRLTKLNSTWEPYFVVGIFCALINGLQ 168
```

```
Db 628 RGGVFNVLTLQVETSKDTEEDLETHIYEKK-----IPVTHSNLVRKSSNTIK 680
QY 117 A--POGDRKLGITKE-DLNENPVVPSFWRLTKLNSTWEPYFVVGIFCALINGLQAFSI 173
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 681 SKYPETEDKEVDEEKKKEGPPPVFFKVMKLNKPEPVYFVVGIVICAMINGATOPAFAI 740
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 174 IFSRIIGITFRDEDPETK-ROIQNMFSVLFLVGLIISITFFLQGGFTFGKAGEILTKRLR 232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 741 IFSRIIGVFA--GPPVSRSESSMYSLFLALGGVSFTFFLQGGFTFGKAGEILTKRLR 797
QY 233 YMVFRMLRDQSVFDDLNKNTGALTTRRLANDAAQVKGATGSRILAVITONIANLGTGIII 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 798 LGSFKSLMRQIEGWFDKSNKSTGALTTRLATDASQVQGTGTRLLAQNANVANGTAIII 857
QY 293 SLIYGWLTLLLLAIVPIIAIAGVVENKMLSGOALKDKKELEGAGKIATEAIENFTVVS 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 858 SFTIYGWLTLLLLAIVPIIAAAGLVEMKPFAGHAKKDKKELEKAGKISTDAVLNIRTVVS 917
QY 353 LTOEOKFEHMYAOSLQ 368
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 918 LTRERKFEAMYKSL 933

RESULT 10
Q9W693 PRELIMINARY; PRT; 851 AA.
AC Q9W693;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MULTIDRUG RESISTANCE TRANSPORTER HOMOLOG (FRAGMENT).
GN MDR.
OS Fundulus heteroclitus (Killifish) (Mummichog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Fundulidae; Fundulus.
OX NCBI_TaxID=8078;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER, INTESTINE;
RA Cooper P.S., Van Veld P.A., Reece K.S.;
RT "P-glycoprotein related sequences from the mummichog (Fundulus heteroclitus)";
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF099732; AAD23956.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR ATP-binding; Transport.
FT NON_TER 1
SQ SEQUENCE 851 AA; 93660 MW; 8F2DC4464171420A CRC64;

Query Match 57.3%; Score 1052; DB 13; Length 851;
Best Local Similarity 55.1%; Pred. No. 2.9e-80;
Matches 209; Conservative 70; Mismatches 80; Indels 20; Gaps 5;
```

Query Match	56.9%;	Score 1044.5;	DB 13;	Length 817;
Best Local Similarity	55.6%;	Pred. No. 1.2e-79;		
Matches	212;	Conservative	67;	Mismatches 81; Indels 21; Gaps 7;
Qy	1	TSALDTESEAVQVALDKARKGRTTIVIAHRLSTRVNADVIAGFDDGVIVKEGNHDELMK	60	
Db	90	TSALDAESETIVQAALDKVRLGRTTIIVAHRLSTRVNADVIAGFDDGVIVKEGNHDELMK	149	
Qy	61	EKGIVFKLVMTQTRNGNEIDLENATGESKESDALEMSPK-----DSGSSL--IKRRSTRS	114	
Db	150	KKGIVHRLVTTQTFQ-----DVE-----EGKEEELSVDEKSPVDSHSECTPYRERKTRGS	202	
Qy	115	--IHAQGDQRKLGTGKEDLNEN-----VPPVSFWRIKLNSTEWPYVFWGIFCAIINGGL	167	
Db	203	SMTVSEGGKEKEKTESDKDEFEDEIVPPVSFFKVLRLNLPEWPIYILVGTICAIINGAM	262	
Qy	168	QPAFSIIFSRIGIFTRDEDPETKQISNMFSLVLVLGIISFTFFFLQGTFFGKAGEIL	227	
Db	263	QPVFALIFSKIIITVFA-ETDQELVRQRTATFLSMLFAVIGGVSFTVMPFQCFCKSGEIL	321	
Qy	228	TKRLRYVFSRMLRDQVSFFDDLKNTTGALTRLANDAAQVKGAGISRLAVITQNTANIG	287	
Db	322	TLKRLGAFKAMMRDGLGDFNPKNSVGALTRLATDAAQVQGGATGVYRMATLQNEFANMG	381	
Qy	288	TGIIISLIYQWLTLLILAIVPIITIAAGVYEMKMLSGOALKDKKELEGAKIATEAIENP	347	

Db 382 TSVIISFYVYGWELTLLILAVVPWVLGAGVEMKLTGHAVEDEKKLEKAGISTEAIENI 444
Qy 348 RTVSVLTQEOKFFEHMYAOSLQ 368
Db 442 RTVASLTPREKFESLYHENLE 462

```

RESULT 12
Q90235 PRELIMINARY; PRT; 1348 AA.
ID Q90235
AC Q90235;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BILE SALT EXPORT PUMP.
OS Raja erinacea (little skate).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalea; Hynosoalea; Pristiorajea; Batoidae;
OC Rajiformes; Rajidae; Raja.
OC NCBI_TaxID=7782;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21340059; PubMed=11447010;
RA Cai S.Y., Wang L., Ballatori N., Boyer J.L.;
RT "Bile salt export pump is highly conserved during vertebrate evolution
RL Am J Physiol. Gastrointest. Liver Physiol. 281:G316-G322(2001).
DR EMBL; AF367243; AAK52958.1; -
SQ SEQUENCE 1348 AA; 148648 MW; 81906F641654CFDD CRC64;
```

Query Match 45.9%; Score 842; DB 13; Length 1348;
Best Local Similarity 43.0%; Pred. No. 3.2e-62;
Matches 171; Conservative 83; Mismatches 108; Indels 36; Gaps

Qy	1	TSALDTESEVVQVALDKARKGRITIVIAHRLSTVRNADVIAGFDGCVIVEKGNHDELK 60
Db	604	TSALDNESAIVQALDKVRFGRITISIAHRLSTVRNVDIIGFEHGRAVERGRHAELLE 663
Qy	61	EKGIFYKLVTMOTRGNETLDENATGESKSDESDALEMPKDSGLIKRRSTRRSIHAPQG 120
Db	664	RKGIYFTLVTLQTQGEALHEKA----RQVNGAIEDCASEK-RQLIRGSSRASVSRTRH 718
Qy	121	QDRK-----LGTK-----EDLNENPVPSVWRILKLNSTE 150
Db	719	QRSRQVSEVLSDLSPGDVASAVRTPFSISLGDKDQVEEESESIEPAPVSRIKLKVNSE 778
Qy	151	WPYFVGTFCAINGGLQPARSIIFSRILGIPTDEDPETKRQISNMFSVLFVLGLTIIF 210
Db	779	WPYMLFGSLGAANGVNPYIALFQSILGTFSIQNEEEKINQI-NAICLFFVVVGVLSVF 837
Qy	211	ITFFLOGTFCKAGEILTFRURYVWFVSMRLRODYSWFDDLKNTTGALLTRLANDAAQVK 270
Db	838	LTLQFLQSFTFAKSGELLFRRLKLCFOAMLROETGCWFDDRKNSPGTLTRLATDASOVQ 897
Qy	271	AIGRS LAVITONIANLGTGIILSYIQWLTLALLAIVPIIAIGVVMKMKSQAULKD 330
Db	898	ATGTQIGMVNSITNIGVSLIIAFVFWKVLTVILCFLPLFALTGALQARMLTGANQDK 957
Qy	331	KELEGAGKIATEATENFTVSVLTQEOKFFEHMYAOSLQ 368
Db	958	EALEAAGOISSELASNINTIAGLAKEMFVOLFEAQLE 995

```

RESULT 13
Q9W6L9 PRELIMINARY; PRT; 765 AA.
ID Q9W6L9
AC Q9W6L9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SISTER OF P-GLYCOPROTEIN (FRAGMENT).
GN SPGP.
OS Fundulus heteroclitus (Killifish) (Mummichog).
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Cyprinodontiformes; Acanthopterygii; Percomorpha; Atherinomorpha;
OX NCBI_TaxID=8078;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Cooper P.S., Van Veld P.A., Reece K.S.;
RT "p-glycoprotein related sequences from the mummichog (Fundulus
heteroclitus).";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF135793; AAD2692.1; -;
DR HSSP; P13569; INBD.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmnm.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 1.
DR SMART; SM00005; ABC_tran; 1.
KW ATP-binding; AAA; 1.
FT NON_TER 1 1
SQ SEQUENCE 765 AA; 83983 MW; 4B74BD1E054B1AF6 CRC54;

Query Match 43.2%; Score 793.5; DB 13; Length 765;
Best Local Similarity 42.7%; Pred. No. 1.9e-58;
Matches 169; Conservative 75; Mismatches 121; Indels 31; Gaps 5;

QY 1 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVKGNHDELMK 60
DB 18 TSALDNESEAVQEQALDKVRGRTTISIALRLSTIKNADIIIVGFHGRAVERGKHNELE 77

QY 61 EKGIVKLVMTQTRGNEIDLENA-TGESKSESDALEMPKDS-----GSSLIKRST 111
DB 78 RKGIVFTLVLTQSGDKALNEKAREMAEGE-EQEPQLNLSRAGSYRASIRQRSRS 136

QY 112 RRSIHAPQGG-----DRKLGTRKEDLNENPVPSFWRILKLNSTEW 151
DB 137 QLSNLIPESVPIAGDLGPRAYSHSHEDYKKAAPAEDEEELVEPAPVARIKYNLP 196

QY 152 PYFVVGIFCAINGLQAPSIIFRSIIGITRDEDPETKROI SNMFSVLFLVGLIGISFI 211
DB 197 PYMLFGLGAAGVGNVPVYSLFSLQILATFA-VTDPEAQRREINGICVFVIVGVSFF 255

QY 212 TFFLQGFTEGKAGILFKRLRYWFRSMRLQDYSWFDLKNITGALTTRLANDAAQVKA 271
DB 256 TOMLQGFYAFKSGELLFRRLRIGFKAMLGQEI GWFDDHRSNPGALTTRLATDASQVGA 315

QY 272 IGSRLAVITONIANLGTGIIISLYGWQLTLLLAIVPIIAIAGVVENKMLSGOALKDKK 331
DB 316 TGSQIGMIVNSLNIAGVIMSFSSNVLLILLCFLPFIALSGGFGQAKMLTGFAKQDKQ 375

QY 332 ELEGAGKIATEINFRTVSVLTQEQKFHYAQLS 367
DB 376 AMESAGVSGEALNNIRTIAGLKGEGSFVEKYEHL 411

RESULT 14
O88331 PRELIMINARY; PRT; 1321 AA.
AC O88331;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE P-GLYCOPROTEIN SISTER.
GN SPGP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;
RA Childs S.J., Yeh R.L., Hui D., Ling V.;
RT "Taxol resistance mediated by the liver-specific Sister gene of P-glycoprotein";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF010597; AAC24753.1; -;
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmnm.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; transport.
SQ SEQUENCE 1321 AA; 146286 MW; 27F67EC366008D0F CRC64;

Query Match 42.6%; Score 782.5; DB 11; Length 1321;
Best Local Similarity 43.5%; Pred. No. 3.4e-57;
Matches 167; Conservative 76; Mismatches 124; Indels 17; Gaps 6;

QY 1 TSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAGFDDGVIVKGNHDELMK 60
DB 586 TSALDNESEAVQEQALNKIQHGHHTIISVAHRLSTVRADVIIGFHHGVAVERGTHEELLE 645

QY 61 EKGIVKLVMTQTRGNEIDLENA-TGESKSESDALEMPKDS-----SPKDSGSLIKRSTR- 113
DB 646 RKGIVFTLVLTQSGDNNAHKETSIMKDATEGTLTERTFSRGSYRDSLRASIRQRSK 705

QY 114 --SIHAP--QGDRKLGTRKED-----LNENPVPSFWRILKLNSTEWVYGFICAIIN 164
DB 706 SLTHTDPPPLAVADHKSSYKDSKDNVDLVVEEPAPVRRILKYNIPENWHILVGSLSA 765

QY 165 GGLQPAFSIIFRSIIGITRDEDPETKROI SNMFSVLFLVGLIGISFITFLQGTFGKAG 224
DB 766 GAVTPYISLLFSQLLGTSLDKQQRSEIHSN-CLFFVILGCVSIFTQLQGYTFKSG 824

QY 225 EILTKRLRYWFRSMRLQDYSWFDLKNITGALTTRLANDAAQVKAIGSRSLAVITONIA 284
DB 825 ELLTKRLKRFKAMLGQDQIGWFDLDRNPNGLTTRLATDASQVQCATGSQVGMVNSPT 884

QY 285 NLGTGIIISLYGWQLTLLLAIVPIIAIAGVVENKMLSGOALKDKKEGAGKIATEAI 344
DB 885 NIITALLIAFFFSWKLSLIIITIFFFLAUSGAVQKMLTGTFASQDKQALEKAGITSEAL 944

QY 345 ENFRTVSVLTQEQKFHYAQLS 368
DB 945 SNIRTVAGIGVEGRFIKAFEVELQ 968

RESULT 15
Q03982 PRELIMINARY; PRT; 538 AA.
AC Q03982;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE P-GLYCOPROTEIN.
GN PGP-1.
OS Cricetus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=91154265; PubMed=1671863;
RA Devine S.E., Hussain A., Davide J.P., Melera P.W.;
RT "Full length and alternatively spliced pgp-1 transcripts in multidrug resistant Chinese hamster lung cells";
RT J. Biol. Chem. 266:4545-4555(1991).

CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

DR EMBL; M59252; AAA37003.1; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmam.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 538 AA; 59432 MW; 09A82E1818AE09C3 CRC64;

Query Match	38.5%;	Score 706;	DB 11;	Length 538;
Best Local Similarity	94.1%;	Pred. No. 2.9e-51;		
Matches 143;	Conservative 6;	Mismatches 3;	Indels 0;	Gaps 0;

QY	217	GFTFGKAGEILTKRLRYMVFRSMLRQDVSWFDDLKNTTGTALTRLANDAAQVKGAGSRL	276
Db	33	GFTFGKAGEILTKRLRYMVFRSMLRQDVSWFDDNKNTTGTALTRLANDAGQVKGATGARL	92

QY	277	AVITONIANLGTGIIISLIYGHQLTLLLAIVPIIAIAGVVMKMLSGQALKDKKKELEGA	336
Db	93	AVITONIANLGTGIIISLIYGHQLTLLLAIVPIIAIAGVVMKMLSGQALKDKKKELEGS	152

QY	337	GKIATEAIENFRVTVSLTQEQKFEHMYAQSLLQ	368
Db	153	GKIATEAIENFRVTVSLTREQKFENMYAQSLLQ	184

Search completed: November 6, 2002, 18:42:57
Job time : 13.4444 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 6, 2002, 19:26:40 ; Search time 640.9 Seconds
(without alignments)
7749.846 Million cell updates

Title: US-09-672-725C-6

Perfect score: 1836

Sequence: 1 TSALDTESEAVVQVALDKAR.....TVVSLTQEQKFEHMYAQS LQ 368

Scoring table: BLOSUM62

22000002	Xgapop	10.0	,	Xgapext	0.5
	Ygapop	10.0	,	Ygapext	0.5
	Fgapop	6.0	,	Fgapext	7.0
	Delop	6.0	,	Delext	7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

100% Processing: Minimum Match 0%
Maximum Match 100%

Maximum March 1908
Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+ p2n.model -DEV=x1h
```

```

model frame = pmr; model dev = xth
-Q=/cgn2_1/USPTO_spool/US09672725/runat_04112002_124342_18001/app query.fasta 1.6435

```

```
-DB=EST -QFMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
```

```
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -L
```

```
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
```

```
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
```

```
-USER=US09672725_@CGN_1_1_7365_@runat_04112002_124342_18001 -NCPU=6 -ICPU=3
```

```
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
```

```
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
```

- YGAPOP=10 - YGAPEXT=0.5 - DELOP=6 - DELEXT=7

Database :

```
EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hc: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vrt: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description
		Score	Match	Length		
1	741	40.4	699	9	BB667773	BB667773 BB667773
2	711	38.7	822	10	B1332761	B1332761 602979574

ALIGNMENTS

RESULT 1	BB667773	699 bp	linear	EST 24-OCT-2001
LOCUS	BB667773	RIKEN full-length enriched, adult male liver tumor Mus		
DEFINITION	musculus cDNA clone C73004OC05 3', mRNA sequence.			
ACCESSION	BB667773			
VERSION	BB667773.1	GI:16399222		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 699)			
	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.			
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Yoshitake Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN)			


```

source
1. .822
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5132490"
/clone_lib="NCI_CGAP_Li9"
/lab_host="DH10B (11 phage-resistant)"
/notes="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 227 a 182 c 200 g 213 t
ORIGIN

Alignment Scores:
Pred. No.: 1.87e-70 Length: 822
Score: 711.00 Matches: 170
Percent Similarity: 76.25% Conservative: 29
Best Local Similarity: 65.13% Mismatches: 46
Query Match: 38.73% Indels: 22
DB: 10 Gaps: 2

US-09-672-725C-6 (1-368) x BI332761 (1-822)

QY 108 argArgSerThrArgArgSerIleHisAlaProGlnGlnAspArgLysLeuGlyThr 127
Db 10 CGTCGACACCGCGCTCGAAAC----- 30
QY 128 LysGluAspLeuAsnGluAsnValPro-ProValSerPheTrpArgIleLeuLysLeuAs 147
Db 31 ---AATGAACGTGATGCAACAGTCCACACGAGTCTTGCGTGAAGGTCTTAAACTGAA 87
QY 147 nSerThrGluTrpProTyrPheValValGlyIlePheCysAlaIleHisAsnGlyGlyLe 167
Db 88 TAAACACAGATGGCCCTACTT-GTGGTGGACACAGCTGTGCCATTCGCCATGGAGCCCT 146
QY 167 uGlnProAlaPheSerIlelePheSerArgIleIleGlyIlePheThrArgAspGluAs 187
Db 147 CCAGCCGGCTTTCATCATCTCTGTAGATAGTATCTTGGCCCT-GGGGATGA 205
QY 187 pProGluThrLysArgGlnIleSerAsnMetPheSerValLeuPheLeuValLeuGlyI 207
Db 206 C---GCAGTGNAGCAGCAAGGTGTACAGTGTCTCCCTGGTCTCTTGGGCTTAGGAGT 262
QY 207 eIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLysAlaGlyGlyIleLe 227
Db 263 CCTCTCTCTTACTTCTCTCTCAGGCGTTCACGTT-GGGAAGCTGGAGAGATCCT 321
QY 227 uThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGlnAspValSerTrpPh 247
Db 322 CACCACAGGCTCGGTCCTCGCTTAAAGCGATGCTTAAGCGAGGACATGAGCTGGT 381
QY 247 eAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaAlaG 267
Db 382 TGATCATCAAAAACAGTACTGGAGCATTTCTACAAGACTTCGCCACAGATGCTGCCCA 441
QY 267 nValLysGlyAlaIleleSerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuG 287
Db 442 AGTCCAAGGACCCACGGGAACAGGTGGCTTAATGTGCACAGACACAGCCACCTTGG 501
QY 287 yThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAlaI 307
Db 502 AACCGTATTATTATATCATTTATTTACGGTTGGCAACTGACACTTCTGCTTATCGGT 561
QY 307 eValProIleIleAlaIleAlaGlyValValGluMetLysMetLeuSerGlyGlnAlaLe 327
Db 562 TGTTCATTCTTCTGTAGAGGAAT-GTTGAAATGAAATGTGGCTGGCAATGCGCAA 620
QY 327 uLysAspLysLysGluLeuGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPh 347
Db 621 GAGAGATAAAAGAAATGGAAC-TGTGGAAGATGTA-ACAGAGGCAATAGAAAATAT 678
QY 347 eArgThrValSerLeuThrGlnGlnLysPheGluHisMetTyrAlaGlnSerLe 367

```

```
Db 181 TTCACTATGTTGGCAACTAACACTGTTACTCTTAGCAATTTGTTACCCATCATTTGCAATA 240
|||||
Qy 314 AlaGlyValValGluMetLysMetLysSerGlyGlnAlaLeuLysAspLysLysGluLeu 333
|||||
Db 241 GCAGGAGTTGTTGAATGAAATGTTGCTGTGACACAGCACTGAAAGATAAGAAAGAACTA 300
|||||
Qy 334 GluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThrValValSerLeu 353
|||||
Db 301 GRAGGTGCTGGGAAGATCCCTACTGAAGCAATAGAAAACCTCCGAACCGTTGTTCTTTG 360
|||||
Qy 354 ThrGlnGluGlnLysPheGluHisMetTyrAlaGlnSerLeuGln 368
|||||
Db 361 ACTCAGGACGACAGAGTTTGAACATATGATGCTCAGAGTTTGCAG 405
|||||

RESULT 4
BF236471 952 bp mRNA linear EST 14-NOV-2000
LOCUS 602025733F1 NCI_CGAP_L19 Mus musculus cdna clone IMAGE:4160913 5',
DEFINITION mRNA sequence.
ACCESSION BF236471
VERSION BF236471.1 GI:111150140
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 952)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9441 row: a column: 10
High quality sequence stop: 651.
Location/Qualifiers
1..952
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_L19"
/lab_host="DHI08 (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 259 a 223 c 259 g 211 t
ORIGIN
source

FEATURES
source
1..951
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4358135"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DHI08 (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies."

Db 81 GTCTCTCCTCTTTACTTCTTCCCTTACGGGCTTCCAGGCTTCCAGGATC 139
|||||
Qy 227 LeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGlnAspValSerTrrp 246
|||||
Db 140 CTCACACAGGCTCCGGTCCATGGCCTTTAAACGCGATGCTAAGGCAGGACATGAGCTGG 199
|||||
Qy 247 PheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaAla 266
|||||
Db 200 TTTGATGATCATATAAAACAGTACTGAGACACTTTTCTACAAGACTCGCCACAGATGTCGG 259
|||||
Qy 267 GlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeu 286
|||||
Db 260 CAAGTCCAAGGAGCCACGGGAACAGGTTGGCTTTAATGCCACAGACACAGCAACACCTT 319
|||||
Qy 287 GlyThrGlyIleIleIleSerLeuIleTyrGlyTrrpGlnLeuThrLeuLeuLeuAla 306
|||||
Db 320 GGAACCGGTATATTATATCATTTTACGGTTGGCAACTGACACTTCTGCTGTTATCG 379
|||||
Qy 307 IleValProIleIleAlaIleAlaGlyValValGlu-MetLysMetLeuSerGlyGlnAl 326
|||||
Db 380 GTTGTTCATTCATTGCTGTAGCAGGAATTTGTAACATGAAATGTGGCTGGCAATGC 439
|||||
Qy 326 aLeuLysAspLys-LysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGluA 346
|||||
Db 440 CAAGAGAGATAAACACCGGAATGGAAGCTTCTCGGAAGATGCAACAGAGCAATAGACA 499
|||||
Qy 346 sn-PheArgThrValValSerLeuThrGlnGlu-----GlnLys 358
|||||
Db 500 ACAATATTCGAACCTGTATATCCTTGACCCAAAGAAACCTTGAGTCAAGTATGTGAAAAA 559
|||||
Qy 359 PheGluHisMetTyrAlaGlnSerLeuGln 368
|||||
Db 560 TTGCATGACCTTACAGGAATTCGGTGGCG 589
|||||

RESULT 5
BF969062 851 bp mRNA linear EST 22-JAN-2001
LOCUS 602269993F1 NIH_MGC_84 Homo sapiens cdna clone IMAGE:4358135 5',
DEFINITION mRNA sequence.
ACCESSION BF969062
VERSION BF969062.1 GI:12336277
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 851)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9995 row: j column: 24
High quality sequence stop: 612.
Location/Qualifiers
1..851
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4358135"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DHI08 (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies."

FEATURES
source
1..851
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4358135"
/clone_lib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DHI08 (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies."

Qy 187 AspProGluThrLysArgGlnIleSerAsnMetPheSerValLeuPheLeuValLeuGly 206
|||||
Db 21 GATGACGACGTGAAGCAAGAGTGTAAACATGCTCCCTGGCTCTTCTGGGCTTAGGA 80
|||||
Qy 207 IleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLysAlaGlyGluIle 226
|||||
```

Note: this is a NIH_MGC Library."

BASE COUNT 243 a 172 c 214 g 222 t
ORIGIN

Alignment Scores:

Pred. No.: 6,86e-45 Length: 851
Score: 487.00 Matches: 101
Percent Similarity: 99.02% Conservative: 0
Best Local Similarity: 99.02% Mismatches: 1
Query Match: 26.53% Indels: 0
DB: 10 Gaps: 0

US-09-672-725c-6 (1-368) x BF696062 (1-851)

QY 267 GlnValLysGlyAlaTleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeu 286
|||||
Db 3 CAAGTTAAGGGCTATAGTTCCAGGCTTGTCTTAATACCCAGAAATATAGCAATCTT 62
QY 287 GlyThrGlyIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAla 306
|||||
Db 63 GGCAGAGGAATAATTATATCTTCATCTATGCTTGCAACTAACACTGTTACTCTTAGCA 122
QY 307 IleValProIleIleAlaIleAlaGlyValValGluMetLysMetLysSerGlyGlnAla 326
|||||
Db 123 ATGTACCCATCATTCATTCATCTATGCTTGCAACTAACACTGTTACTCTTAGCA 182
QY 327 LeuLysAspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsn 346
|||||
Db 183 CTGAAGATAGAAAGAACTAGAGGCTCTGGGAAGATCGCTACTGAAGCAATAGAAAC 242
QY 347 PheArgThrValValSerLeuThrGlnGluGlnLysPheGluHisMetTyrAlaGlnSer 366
|||||
Db 243 TTCCGAACCGTGTGTTCTTGACTCAGGAGCAGAGTTTGAACATATATGTCCTCAGAGT 302
QY 367 LeuGln 368
|||||
Db 303 TTGCAG 308

RESULT 6
AW421821
LOCUS
DEFINITION
186f02.y1 Sugano Kawakami zebrafish DR1 Danio rerio cDNA clone
264539 5' similar to TR:093437 093437 ABC TRANSPORTER PROTEIN. ;
mRNA sequence.

AW421821
AW421821.1 GI:6949753

EST.

SOURCE

zebrafish.

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.

1 (bases 1 to 666)

Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S.,
Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T.,
Martin, J., Pape, D., Steptoe, M., Underwood, K., Theising, B., Ritter
E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU Zebrafish EST Project 1999

Unpublished (1999)

Other_ESTs: fi86f02.x1

Contact: S.L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by Washington University Genome Sequencing Center

Seq primer: T3 ET from Amersham

High quality sequence stop: 469.

Location/Qualifiers

FEATURES

source

1..666
/organism="Danio rerio"

/strain="AB"
/db_xref="taxon:7955"
/clone_lib="264539"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGCTG);
Site_2: DraIII (CACCATGT); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TCTTGGCCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGCTG, 3' site
CACCATGT). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTCTGCTCTTAAAGCTGGC and 3' end
primer CGACCTGCAGCTCGAGACA."

BASE COUNT 181 a 150 c 164 g 170 t

ORIGIN

Alignment Scores:

Pred. No.: 2.14e-43 Length: 666
Score: 472.50 Matches: 91
Percent Similarity: 66.50% Conservative: 42
Best Local Similarity: 45.50% Mismatches: 64
Query Match: 25.74% Indels: 3
DB: 9 Gaps: 2

US-09-672-725c-6 (1-368) x AW421821 (1-666)

QY 106 IleLysArgSerThrArgArgSerIleHisAlaProGlnGlyGlnAspArgLysLeu 125
:::|||||
Db 76 CTGAACGTTCCAGTAAGAGCGCTCTCTCCAGAAAGAGAGCAAGAAATCCACAAA--- 132
QY 126 GlyThrLysGluAspLeuAsnGluAsnValProValSerPheTrpArgIleLeuLys 145
|||||
Db 133 ---GACAAAAGGCAAAAAGAGAAATCTCTGAAATGCCATTCATAAATCCCTGGCA 189
QY 146 LeuAsnSerThrGluTrpProTyrPheValValIlePheCysAlaIleIleAsnGly 165
:::|||||
Db 190 CTGACAAACCAACCACTGGCTTACCTGGTGGGCGACCTCCCGACCTCGGGTGGGA 249
QY 166 GlyLeuGlnProAlaPheSerIleIlePheSerArgIleGlyIlePheThrArgasp 185
:::|||||
Db 250 GCAGTGTATCCCTCGCTGGCCATCTCTTGCACAAATCATTTGCTATTGCA---GAA 306
QY 186 GluAspProGluThrLysArgGlnIleSerAsnMetPheSerValLeuPheLeuValLeu 205
|||||
Db 307 CCGGACCCCTGAGTGAAGCGGCGAGAACTATGATGTTCTCATTATACCTGCTAATT 366
QY 206 GlyIleIleSerPheIleThrPhePheLeuGlnClyPheThrPheGlyLysAlaGlyGlu 225
|||||
Db 367 GGAGTTGTGGCTTTCTCATTATTTCTCCAGAGCTTCATGTTGGTAAATCCCGGTAG 426
QY 226 IleLeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGlnAspValSer 245
:::|||||
Db 427 CTCCTAACCATGAGATTGAGAGCCAGGCTTTCAGGCAATTTGTAGACAGAGATTGGC 486
QY 246 TrpPheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArgLeuAlaAsnAspAla 265
|||||
Db 487 TGGTTTGTATGACAAACAACTAGTGGGAATCTCCACCCTAAATATGCCACAGATGCG 546
QY 266 AlaGlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsn 285
:::|||||
Db 547 TCTCTGTTAAAGTCTGCTCATCATCCACATTANGACTGGCCGCCACCACTATCTGTGCT 606
QY 286 LeuGlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeu 305
|||||
Db 607 CTACTCATTTGATGTCGTGGCTTCGACATCTCTGTGGCAGCTCACATTTACTGATCTC 666

```

RESULT 7
AV287606
LOCUS
DEFINITION
AV287606 RIKEN full-length enriched, 11 days pregnant adult female
ovary and uterus Mus musculus cDNA clone 5031438E12, mRNA sequence.
ACCESSION
AV287606
VERSION
AV287606.2 GI:15405807
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 688)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ono,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Nov 9, 1999 this sequence version replaced gi:6298110.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
,K., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
Ishii,Y. and Hayashizaki,Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
Func. Genomics 2 pre, L72-L86 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers
1. 688
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5031438E12"
/clone_lib="RIKEN full-length enriched, 11 days pregnant
adult female ovary and uterus"
/sex="female"
/tissue_type="ovary and uterus"
/dev_stage="11 days pregnant, adult"
FEATURES
source

```

```

/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTTAATAATATCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from lambda FIC I. Cloning sites, 5' end: Sali; 3' end:
BamHI."
BASE COUNT 193 a 132 c 155 g 208 t
ORIGIN
Alignment Scores: 1.08e-40 Length: 688
Pred. No.: 449.00 Matches: 84
Score: 89.92% Conservat: 23
Percent Similarity: 70.59% Mismatches: 12
Best Local Similarity: 24.46% Indels: 0
Query Match: 9 Gaps: 0
DB:
US-09-672-725c-6 (1-368) x AV287606 (1-688)
QY 250 LeuLysAsnThrThrArgLeuAlaLeuThrArgLeuAlaAsnAlaGlnValLys 269
Db 11 ATAAACAGCGCATTTGGTTCATTGACCCAGGCTCGCCAGTGAAGTTTGTAGTGA 70
QY 270 GlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeuGlyThrGly 289
Db 71 GGGGGGATGGGCGCGGCTGCTGTGTACCCAGAGATGTAGCAAACTCGGGACAGGA 130
QY 290 IleIleIleSerLeuIleTyrGlyTyrGlnLeuThrLeuLeuLeuAlaIleValPro 309
Db 131 GTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 190
QY 310 IleIleAlaIleAlaGlyValValGluMetLysMetLysMetLysGlyGlnAlaLeuLysasp 329
Db 191 CTCAATTGTATTGGCGGAATTATTGAAATGAAGCTGTTCTGCGCAAGCCTTGAAGGAC 250
QY 330 LysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArgThr 349
Db 251 AGAAACACGCTTGAGATCTCTCGGGAAGATTGCTACAGAGCAATTTGAAACTTCGCGACT 310
QY 350 ValValSerLeuThrGlnGluGlnLysPheGluHisMetTyrAlaGlnSerLeuGln 368
Db 311 ATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 367
RESULT 8
BB861025 463 bp mRNA linear EST 26-NOV-2001
LOCUS BB861025 RIKEN full-length enriched, pooled cell lines Mus musculus
DEFINITION CDNA clone G430017B02 5', mRNA sequence.
ACCESSION BB861025
VERSION BB861025.1 GI:17102479
KEYWORDS EST.
SOURCE house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 463)
Akinura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.

```


strand cDNA was primed with a Not I - oligo(dT)15 primer [5'pCAGTAGTCAGATCGGAGCGCCCTTTT3']; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRP, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss) adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control.

BASE COUNT 105 a 120 c 133 g 160 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 2,99e-40 Length: 519
Score: 443.50 Matches: 78
Percent Similarity: 71.60% Conservative: 43
Best Local Similarity: 46.15% Mismatches: 47
Query Match: 24.16% Indels: 1
DB: 9 Gaps: 1

US-09-672-725C-6 (1-368) x AN018885 (1-519)

Qy 147 AsnSerThrGluTrpProTyrPheValGlyIlePheCysAlaIleLeuAsnGlyGly 166
Db 13 AACGCTCTGAATGCGCCCTACATGTTTGGGAAGCATGGAGCGCTCGAATGAGGG 72
Qy 167 LeuGlnProAlaPheSerIlePheSerArgIleIleGlyIlePheThrArgSpGlu 186
Db 73 GTAAACCCAGCTACCTCTCTCTTTTCAGTCAAACTTGGCGACATTTTCGATGCGAGAT 132
Qy 187 AspProGluThrLysArgGlnIleSerAsnMetPheSerValLeuPheLeuValLeuGly 206
Db 133 CCGTGTGAACACAGGAGGAGAT---AATGGCATCTGCTGTCTTCTAGTAGTTGT 189
Qy 207 IleIleSerPheIleThrPhePheLeuGlnGlyPheThrPheGlyLysAlaGlyGluIle 226
Db 190 TTGGTTCTTTTTCACACAAATGTTACAGGGCTATGTCATTTCTAAATCCGCGCAGCTG 249
Qy 227 LeuThrLysArgLeuArgTyrMetValPheArgSerMetLeuArgGlnAspValSerTrp 246
Db 250 CTCACCTGATAGTTAGGCGCTCGGCTTTCACGCAATGTTGGCGCAGGATCGCTGG 309
Qy 247 PheAspAspLeuLysAsnThrThrGlyAlaLeuThrThrArgLeuAlaAsnAspAlaAla 266
Db 310 TTTGATGATCGCAAAACAGCCCTGTGTGTTTAAACACAGCACTGGCTACAGATGCTCG 369
Qy 267 GlnValLysGlyAlaIleGlySerArgLeuAlaValIleThrGlnAsnIleAlaAsnLeu 286
Db 370 CAAGTCAGGAGGCCACAGGTTCTCAGATTGGCATGATTGTGAATTTCTCTGACCAACATC 429
Qy 287 GlyThrGlyIleIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAla 306
Db 430 GGTNGTCTGTTATCATCTCTTTTACTTCAGCTGGAAGCTCACTGCTGCTGCTGCTG 489
Qy 307 IleValProIleIleAlaIleAlaGly 315
Db 490 TTCCTTCCATTCTGGCTCTTCTGGA 516

RESULT 10
AZ692194/c 848 bp DNA linear GSS 14-DEC-2000
LOCUS ENT001194 Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION AZ692194
VERSION AZ692194.1 GI:11829340

KEYWORDS
SOURCE

Entamoeba histolytica.
Entamoeba histolytica

ORGANISM

Eukaryota; Entamoebidae; Entamoeba.

REFERENCE

1 (bases 1 to 848)

AUTHORS

Loftus,B., Van Aken,S. and Fraser,C.

TITLE

Determination of clone end sequences from Entamoeba histolytica

JOURNAL

HM1:IMSS sheared DNA library

COMMENT

Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjl@loftus.tigr.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared

DNA library

Seq primer: M13-Reverse

Class: Shotgun

High quality sequence start: 35

High quality sequence stop: 685.

Location/Qualifiers

1..848

/organism="Entamoeba histolytica"

/strain="HM1:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOSt; Site 1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1999).

BASE COUNT 272 a 157 c 105 g 314 t

ORIGIN

Alignment Scores:

Pred. No.: 7.76e-32 Length: 848

Score: 372.50 Matches: 95

Percent Similarity: 45.78% Conservative: 46

Best Local Similarity: 30.84% Mismatches: 85

Query Match: 20.29% Indels: 82

DB: 12 Gaps: 4

US-09-672-725C-6 (1-368) x AZ692194 (1-848)

Qy 1 ThrSerAlaLeuAspThrGluSer-GluAlaValValGlnValAlaLeuAspLysAlaAa 20

Db 812 ACTTCTGCATTAGATTATTAATCAAGAAGAATGTCCAAAGACATTAGTAGGTGTC 753

Qy 20 glySGlyArgThrThrIleAlaHisArgLeuSerThrValArgAsnAlaAspVa 40

Db 752 AGCTGGAAGACATCAGTTATTATTCACATAGATTATCTACAATTAGATGACATAG 693

Qy 40 IlleAlaGlyPheAspGlyValIleValGlyGlyAsnHisAspGluLeuMetLy 60

Db 692 AATCTTCTGTGACCATGGACAAAGTTGTTGAAGATGGAATATTATCAACATTAAATGGA 633

Qy 60 sGlyLysGlyIleTyrPheLysLeuValThrMetGlnThrArgGlyAsnGluIleAspLe 80

Db 632 AAACACAGATTATTTTAACTCGTTAAATAATCAAGAAATGGGAAGAAA----- 581

Qy 80 uGluAsnAlaThrGlyGluSerLysSerGluSerAspAlaLeuGluMetSerProLysAs 100

Db 581 ----- 581

Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES
source
1. .605
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="ELOV006G02"
/tissue_lib="Bos taurus ovary fetus"
/tissue_type="ovary"
/dev_stage="fetus"
/lab_host="DH108"
/note="Vector: pZLL; Site_1: Sall; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 158 a 128 c 148 g 171 t
ORIGIN

Alignment Scores:
Pred. No.: 1.18e-31 Length: 605
Score: 369.00 Matches: 74
Percent Similarity: 100.00% Conservative: 6
Best Local Similarity: 92.50% Mismatches: 0
Query Match: 20.10% Indels: 0
DB: Gaps: 0

US-09-672-725C-6 (1-368) x AV616675 (1-605)

QY 289 GlytIleIleSerLeuIleTyrGlyTrpGlnLeuThrLeuLeuLeuAlaIleVal 308
|||||
Db 12 GGCATTATTATGCCCTCATCTATGGCTGGCAGTTACACTTCTCCGTGAGCAATTGTA 71
|||||

QY 309 ProIleLeaIleAlaGlyValValGluMetLysMetLeuSerGlyGlnAlaLeuLys 328
|||||
Db 72 CCCATCATTCAGTAGCAGGAGTTATTGAAATGAAATGTTCTCGACAGCCCTGAAA 131
|||||

QY 329 AspLysLysGluLeuGluGlyAlaGlyLysIleAlaThrGluAlaIleGluAsnPheArg 348
|||||
Db 132 GATAAGAAAGAACTGGGAAGTGTGGGAAGATTGCCACTGAAGCGATPAGAAAATCCGG 191
|||||

QY 349 ThrValValSerLeuThrGlnGluGlnLysPheGluHisMetYrAlaGlnSerLeuGln 368
|||||
Db 192 ACTGTTCTTTTGACTCGGAGGAGAGTTGAATATATATATGATGCCAGAGTTGCAG 251
|||||

RESULT 12
BH459648/c

LOCUS
BH459648 836 bp DNA linear GSS 13-DEC-20
BOGSB90TF BOGS Brassica oleracea genomic clone BOGSB90, DNA
sequence.

ACCESSION
BH459648
VERSION
BH459648.1 GI:17649393
KEYWORDS
GSS.

SOURCE
Brassica oleracea.
ORGANISM
Brassica oleracea.

REFERENCE
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

AUTHORS
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE
Whole genome shotgun sequencing of Brassica oleracea
JOURNAL
Unpublished (2001)
COMMENT
Other_GSSs: BOGSB90TF
Contact: Chris Town
TIGR

7121 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES
source
1. .836
/organism="Brassica oleracea"
/strain="T01000DH3"

Db 32 ATCTTGTTCCGATT 18
|||||

Search completed: November 7, 2002, 07:28:38
Job time : 651.9 secs

